Implementation of Deep Learning based Insurance Premium Prediction System

A thesis submitted by

Ayesha Faheem (2021-BSE-005)

Supervisor	**Co-Supervisor**
Dr. Irum Matloob	Engr. Sumaira Shaukat

Submitted in partial fulfillment of the requirements for the degree of
Bachelor of Software Engineering

Faculty of Engineering

Department of Software Engineering
Fatima Jinnah Women University The Mall Road, Rawalpindi

May 2025

DEPARTMENT OF SOFTWARE ENGINEERING

Certificate

This is to certify that **"Implementation of Deep Learning based Insurance Premium Prediction System"** is submitted in partial fulfillment of the requirement for the degree of Bachelor of Software Engineering by the student:

Ayesha Faheem (2021-BSE-005)

Supervisor Dr. Irum Matloob	**Co-Supervisor** Engr. Sumaira Shaukat
External Examiner	**Chairman**

Date: ____________________

Acknowledgment

This research represents a major turning point in our academic journey and would not have been possible without the countless blessings of Allah Almighty. We will always be grateful to Almighty for giving us the courage, patience, and wisdom to carry out and finish this thesis effectively. His guidance has been our compass through the most difficult phases of this work.

The authors express their sincere gratitude to Dr.Irum Matloob, our supervisor, and Engr. Sumaira Shaukat, our co-supervisor, whose outstanding mentorship, insightful feedback, and constant support were essential to our progress. Their wisdom, support and trust enabled us to explore our ideas confidently. We are especially grateful for their assistance with the dataset.

We also express our sincere gratitude to our loving parents. Their unwavering belief in us, unending support, and emotional strength carried us through every challenge. Their sacrifices and constant support have been the foundation of our success.

Dedication

We dedicate this thesis to our beloved parents, whose support, love, and prayers have been our greatest strength. Their sacrifices, encouragement and belief in our potential have guided us through every challenge. We also dedicate this thesis to our respected mentors and faculty, who have guided and encouraged us throughout this journey. Above all, we are grateful to Almighty Allah for his endless blessings on us.

Contents

List of Tables

List of Figures

List of abbreviations

RNN	Recurrent Neural Network
LSTM	Long-Short Term Memory
GAN	Generative Adversarial Network
AI	Artificial intelligence
KNN	K-Nearest Neighbors
GLM	Generalized Linear Model
GBM	Gradient Boosting Machine
FFNN	Feedforward Neural Network
CANN	Convolutional Artificial Neural Network
FCM	Fuzzy C-Means
DNN	Deep Neural Network
SVM	Support Vector Machine

List of Symbols

$\phi_1, \phi_2, \ldots, \phi_7$	Input features representing patient-related attributes
η	Learning rate used for weight updates
$\sigma(\cdot)$	Activation function (ReLU, tanh)
y_t	Predicted output at time step t
$\sum(y_{Tt} - \hat{y}_t)^2$	Mean Squared Error (MSE) loss function
$\sigma(o\ )$	Softmax function to compute category probabilities
O_t	Output before activation
S_i	Service sequence for a patient (list of unique service IDs)
$\tilde{S}_i$	Processed sequence after padding/truncation
h_t	Hidden state of the LSTM at time step t
θ	Set of trainable parameters for the LSTM model
τ	Anomaly detection Threshold
d	unique doctor identifier
S_d	Sequence of services provided by doctor d
L	Fixed sequence length
L	Binary cross-entropy loss function
θ	Model parameters
η	Learning rate

Abstract

Traditional health insurance models lack flexibility in insurance coverage, these models suggest fixed insurance policies and plans for enterprises. Due to this, the insured employee is not adequately covered and bears additional expenses from own sources over and above the claim amount. Enterprises pay equal premium amounts for all employees even for those who infrequently utilize healthcare insurance services. This results in inadequate cost distribution. The manual and static evaluations and verification methods are used for processing insurance claims which may not fully mitigate the risks of entertaining fraudulent claims and ultimately enterprises and insurance companies may bear additional financial burdens. A transformative approach is needed, which can integrate insurance packages and anomaly detection techniques for prevention of fraud, risk evaluation and claim optimization. The proposed framework introduces an efficient way to predict health insurance premiums and generate customizable insurance plans using Recurrent Neural Networks (RNN) and clustering techniques. The proposed methodology attempts to predict the insurance premium amounts and splits patients into three categories of risk- Low, Moderate, and High Risk- depending upon how they utilize healthcare services. These classifications help to create dynamic and need-based insurance policies instead of the rigid designation-based policies. Moreover, to prevent and to monitor fraudulent practices in healthcare, we have applied LSTM-based Anomaly Transformer and Generative adversarial networks and compared both models results. Experimental results show that Anomaly transformer is demonstrating better results for our case.

The proposed framework efficiently detects anomalies in employee and physician behavior, with a validation loss of 0.0917 and 96.4% validation accuracy. The research demonstrates that the proposed AI driven framework can perform both coverage optimization using multiple patient data sets, as well as decision support in healthcare systems.

Chapter 1

Introduction

1.1 Background

Healthcare costs continue to rise, which makes health services unaffordable for both individuals and enterprises. To manage these financial burdens, many individuals choose to purchase health insurance. However, even if insurers offer coverage under fixed rates, often these policies do not adequately reflect the health needs of individual policy holders. As a result, policyholders can expect to pay substantial out-of-pocket costs for health care services [1]. This financial burden disproportionately affects lower income consumers and individuals with chronic illnesses, forcing them to delay essential treatments or take on medical debt. A major flaw in traditional health insurance models is how premiums are calculated. Instead of being allocated based on an individual's actual healthcare needs, many insurance companies calculate premiums by assigning coverage based on some general factors such as job designation, age group, or company policies. This outdated approach results in an excessive imbalance of coverage, where some employees get too much coverage that they do not fully utilize, while others face inadequate insurance coverage that does not adequately cover their essential medical needs [2]. Such misallocations not only results in unintended costs for individuals but also lead to unnecessary financial burdens for employers, who often pay a subsidy to employee insurance plans without assuring that they align with actual healthcare utilization. An employee under insurance may have very high out-of-pocket costs and end up experiencing financial stress and medical debt, while an employee over insurance receives more complete coverage that they actually do not need, increasing the cost of the overall insurance plan in overall terms. These inefficiencies create a healthcare system in which healthcare is still unaffordable [3].

1.2 Problem Statement

As healthcare services become costly, getting health insurance coverage becomes a requisite for most. Mostly, companies, firms, and other organizations offer health insurance services in order to aid their workers financially when there is any medical attention required. Unfortunately, the existing models of insurance are unsympathetic and poorly constructed, which results in two main issues:

1. **Enterprise-Level Issue:** Employees of different organizations are offered insurance plans with high premiums by their employers; however, some employees never use the allocated insurance coverage. Organizations have to deal with these costs, which unfortunately add to the business expenditure without providing any value in return.

2. **Employee-Level Issue:** Each insurance plan caters to a specific designation and not the exact healthcare requirement of the employee under that designation. Because of this:

- Some employees have to incur additional out-of-pocket expenditure as their insurance plans do not sufficiently cover their medical needs.
- Some perform fraudulent or anomalous healthcare transactions.
- Others may be over insured, meaning that they have more coverage than they require which results in financial wastage.

The primary objective is to maintain balance between two crucial factors: employer budget and employee need. An imbalance between these two aspects results in stressed-out employees or overworked employees and employees engaging in illegal or fraudulent transactions. The imbalance between these two factors results in overburdened employees or stressful employees as shown in 1.1.

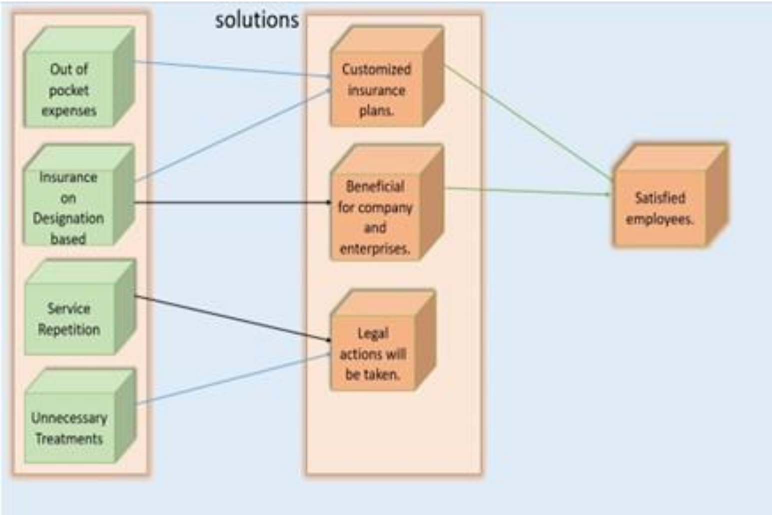

Figure 1.1: Description of problems to be addressed

This type of planning would not offer any flexibility, which is the biggest issue with current models. Existing plans try to preset, target, and then ignore important individual markers like their current health issues, personal health records, lifestyles, and most importantly what you need. If these criteria are missed so much they are leaving employees on medical care packages that just don't have enough options to adequately cover the needs of their healthcare. The solution employs sequence modeling techniques like RNN and LSTM-based Anomaly Transformer to detect unusual or out-of-pattern service usage behaviors, while clustering algorithms such as K-Means and Fuzzy C Means are applied to group individuals with similar healthcare needs, making it possible to create personalized, need-based insurance plans. A system like this would consider non-exhaustive factors like an employee's health record, previous medical costs, and any service risk factors to recommend a customized insurance package. This imbalance would leave businesses with only paying for the minimum required coverage for their employees, while employees would be insured in a way that minimizes their out of-pocket payment for health services. Adopting this kind of intelligent, flexible, and data-driven insurance policy would result in a more equitable and efficient insurance system for both employees and enterprises, translating into better expense management and healthcare service access.

1.3 Aims and Objectives

The primary aim of this study is to develop and design an intelligent need-based premium prediction system that makes use of deep learning techniques to guarantee equitable, personalized and effective healthcare coverage. The proposed system addresses the shortcomings of traditional designation-based premium models, which frequently results in resource wastage and financial imbalance for both employees and organizations. This study aims to optimize premium allocation based on actual healthcare service utilization by implementing a data driven strategy. To achieve study aims, the key
objectives are as follows:

- To gather real-world insurance claim data from a local hospital and perform pre-processing.
- To implement RNN for predicting insurance amount, and future medical service usage.
- To apply LSTM-based Anomaly Transformer for detecting abnormal behavior of doctors.
- To implement K-Means clustering algorithm for generating personalized insurance plans.
- To develop an intelligent, need-based premium prediction system that provides equitable and optimized premium suggestions to enterprises for effective healthcare cost distribution.

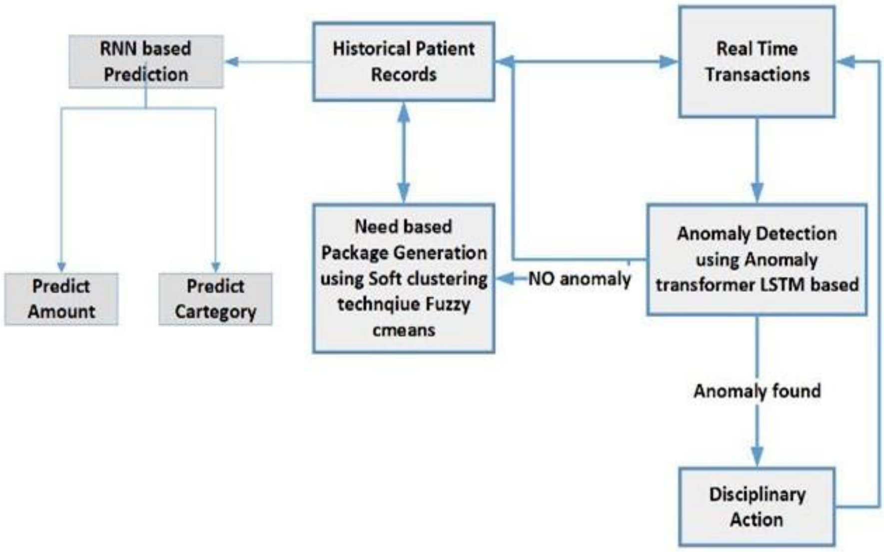

Figure 1.2: System Overview Diagram

1.4 Thesis Layout

Basically, the thesis consists of seven sections.

Section 1: This section outlines the problem related to designation-based insurance premium prediction. It also presents the proposed solution, a need-based insurance premium prediction system using deep learning techniques.

Section 2: This section provides literature review of existing research on insurance premium prediction and fraud detection.

Section 3: This section explains the step-by-step approach followed in model building.

Section 4: In this section, system performance, accuracy, and overall effectiveness are analyzed.

Section 5: Summarizes the key findings, research contributions, and practical impact of the proposed solution. It also discusses potential enhancements and directions for future research.

Chapter 2

Literature Review

Extensive research in healthcare premium prediction has thoroughly investigated numerousmachinelearninganddeeplearningapproachestoenhanceaccuracyandiden tify fraud. Artificial Intelligence (AI) has also been used to optimize premium computation, claim processing and risk evaluation. This section will explore the most recent studies addressing the key findings and limitations, while also identifying potential areas for further improvements.

Kaushik et al. [4] developed an ANN-based model that predicts insurance premium with 92.72% accuracy, but that is built from static factors such as age and BMI, and didn’t include dynamic parameters like health parameters and personalization of risk curve. Vuddanti et al. [5] applied regression and gradient boosting models on 4,968 records dataset, with 94% accuracy, but with only structured data available and no real-time adaptive capability. Similarly, Kulkarni et al. [6] used Gradient Boosting Regression (86.86% accuracy) to predict cost but did not explore deep learning-based methods and adaptation to unstructured data. Jyothsna et al. [7] reported a 87% accuracy in health insurance premium prediction by XGBoost and developed a Telegram chatbot for cost estimation. Their study did not explore deep learning models or user engagement. AbdElminaam et al. in order to estimate medical costs, [8] examined five machine learning models: Although decision trees had the lowest MAPE (3.5%) and were the most accurate, they did not adjust effectively to real-time data. Reddy et al. [9] conducted linear regression and ANOVA analysis on medical expenses using smoking, age, and BMI, concluded that smoking is the most important factor, but also did not take disease history or other medical conditions which affect the insurance costs. Mavundla et al. [10] presented an investigation of health insurance cross-selling using Random Forest, KNN, XGBoost, and Logistic Regression on 1,000,000 customer records, whereby Random Forest achieved 99% accurate model. The study did not account for the external factors like regulation and economic conditions that may impact the insurance pricing and policy adoption. A study from IIST 2024 [11] applies machine learning to premium prediction and finds improvements in accuracy and risk assessment, but challenges with respect to scalability and real-time adaptiveness. Holvoet et al. [12] analyzed neural networks for insurance pricing and tested the performance of GLM, GBM, FFNN, and CANN models. The findings are discussed in detail, and the predictions made by CANN demonstrated significant improvements in predictive accuracy, but the challenges faced by insurance pricing are still significant. Haji Mohammad [13] proposed a data-driven optimization method for insurance premium calculation based on classification models and linear programming. Their approach improved the risk assessment but required extensive preprocessing. Dong and Quan [14] developed an AutoML framework for insurance data analytics which focuses on automated model selection and hyperparameter tuning, but it does not use domain-specific customization for need-based healthcare insurance prediction. Patra et al. [15] analyzes health insurance cost prediction using machine learning techniques by evaluating XGBoost, Lasso, Ridge, and KNN models. While XGBoost performs best, the study does not investigate deep learning techniques and real-time adaptability. Albalawi et al [16] used machine learning to predict healthcare costs using polynomial regression which demonstrates superior accuracy with deep learning methods. However, this study does not explore deep learning methods or real-time adaptability. Kshirsagar et al. [17] developed a machine learning-based framework for transparent health insurance pricing, improving cost predictions by 20% over actuarial models. However, the study lacks real-time adaptability and does not address need-based plan generation. Alam and Prybutok [18] applied machine learning to predict healthcare insurance claims, with XGBoost and Random Forest achieving the highest accuracy. However, the study does not explore deep learning techniques or real-time adaptability. Ul Hassan et al. [19] used ML models for medical insurance cost prediction, with Stochastic Gradient Boosting performing best. However, it lacks deep learning and real-time adaptability.

The deep neural networks were used by Samiuddin et al. [20] to predict health insurance premiums, and the results showed an improvement in accuracy compared to traditional models. However, the study relies on static features and lacks real-time adaptability. Drewe-Boss et al. [21] used 1.4 million health insurance datasets to show that deep learning models outperform ridge regression for cost prediction, capturing meaningful complicated medical interactions. However, interpretability issues and data sparsity remain unresolved. Chu et al. [22] examined deep learning as a cross-selling for health insurance, where TabPFN demonstrated a significant improvement in accuracy (from 52.62% to 63.10%), but dataset imbalance negatively impacted the model's reliability. Patil et al. [19] applied computational intelligence to predict medical insurance cost, achieving higher accuracy but requiring large datasets and extensive preprocessing. In another study, Morid et al. [23] reviewed deep learning techniques for healthcare time series prediction, and the results are quite promising in terms of patient representation and scalability. However, their study described some limitations in cross-setting evaluations and model interpretability, which hinder real-world applications. Recurrent Neural Network for sequence learning are critically reviewed by [24], with applications in finance, healthcare and NLP. The research emphasized RNNs advantages in prediction and sequential learning but also noted challenges in training complexity and vanishing gradients, which limit scalability. Sutskever et al. [25] introduced LSTM-based sequence-to-sequence learning in machine translation and achieved a BLEU score of 34.8. While reversing input sequences, the model improved long-sequence translation accuracy, but reported problems in off-vocabulary words and large number of training data.

Fursov et al. [26] used sequence embeddings and deep learning to detect fraud in health insurance claims with ROC AUC of 0.873 which performs better than traditional models, but lacked interpretability, making real-world implementation challenging. Flaig and Junike [27] used GANs in insurance market risk modeling and found improvements in scenario generation but difficulties in interpretation for regulatory approval. Gomes et al. [28] used autoencoders to detect fraud in insurance claims. However, its focus relates to personalized healthcare premium prediction that is limited to need-based models. The research by [29] uses LSTM-based models to detect fraudulent healthcare claims. Although it is effective, it is based on large, labeled datasets and model's performance is sensitive to imbalanced data which may lead to potential bias in classification of anomalies. The research by [30] proposes a sequence mining-based fraud detection framework for healthcare insurance that analyzes patient visit patterns across specialties to identify anomalies. While effective, the model is very dependent upon predefined rules and historical data which limits its adaptability to new fraud tactics. Another research by [31] proposes a fraud detection framework for government healthcare programs that relies upon predefined patterns which limit its adaptability to new fraud tactics. The research by [32] employed machine learning models to detect insurance fraud, and Decision Trees achieved an accuracy of 79%. However, it lacks deep learning and real-time detection. To identify 19 important fraud signs, Nalluri et al. [33] created machine learning models that included Decision Trees, SVM, Random Forest, and MLP. However, the study lacks deep learning approaches and real-time adaptability. Zhang et al. [34] integrated block chain and deep learning for medical insurance fraud detection, utilizing BERT-LE for classification. However, it relies on predefined rules and lacks adaptability to new fraud patterns. Dey et al. [35] explored AI-driven fraud detection in U.S. healthcare, integrating supervised and unsupervised ML techniques for billing risk assessment. However, the study faces challenges in regulatory compliance, privacy concerns, and model interpretability. The research by Pandya [36] suggests Artificial Neural Network (ANN) for identifying fraudulent activities in healthcare insurance, in contrast to conventional models like Ridge and XGBoost. However, the study is limited by a small dataset and lacks real-time adaptability. For effective health insurance management [37] suggested a block chain and machine learning based architecture that uses random forest for risk classification and ridge regression for Amount prediction. However, the research lacks deep learning based premium modeling and real time claim processing flexibility. The research by Alcantara et al. [38] applies machine learning to optimize healthcare insurance premiums, enhancing risk assessment and cost prediction. However, the study relies on historical data and lacks real-time adaptability.

Srinivasagopalan [39] applied Deep Q-Learning for dynamic healthcare insurance pricing but focused on plan generation and broader deep learning-based risk assessment. Matloob et al. [40] employed machine learning and clustering algorithms to optimize need-based health insurance packages, reducing medical benefit costs by 25%. However, the study lacks real-time adaptive pricing models and does not include deep learning techniques for premium prediction.

Table 2.1: Key research papers on healthcare insurance predictability & fraud detection

Author(s)	**Methodology**	**Key Findings**
Kaushik et al. [4]	ANN-based model	Achieved 92.72% accuracy but lacked dynamic health parameters.
Drewe-Boss et al. [21]	Deep learning vs Ridge Regression	Deep learning captured difficult medical interactions, but struggled to prove model interpretability.
Morid et al. [23]	Deep learning for time-series healthcare prediction	Effective for patient representation but insufficient cross-setting evaluations.
Vuddanti et al. [5]	Regression, Gradient Boosting	Achieved 94% accuracy but no real-time adaptability.
Jyothsna et al. [7]	XGBoost, Telegram chatbot	Achieved 87% accuracy, but deep learning models are not considered.
Fursov et al.[26]	Sequence embeddings, Deep Learning	Achieved ROC AUC 0.873 for fraud detection but no interpretability.
Mavundla et al. [10]	Random Forest, KNN, XGBoost	Random Forest achieved 99% accuracy, but regulatory and economic factors are not taken into account.

Zhang et al. [34]	Block chain, Deep Learning (BERT-LE)	Improved fraud detection but required predefined rules.
Matloob et al. [40]	LSTM-based fraud detection	Effective but sensitive to imbalanced datasets.
Srinivasagopalan [39]	Deep Q-Learning For insurance pricing	Effective but didn't focus on plan generation.

However, even though deep learning and machine learning methods have significantly increased the accuracy of fraud detection and health insurance premium prediction [12], [41], several significant challenges remain to be resolved. The dominant emphasis in existing research is on improving predictive accuracy, while failing to consider the explain ability and fairness of these models. One of the biggest limitations on insurance pricing is the absence of transparent and unbiased decision-making. Most research focuses on deep learning-based fraud detection [14], but the extensive use of reinforcement learning for optimal healthcare insurance pricing has not received much attention [28]. Addressing these gaps can be remedied through more transparent, ethical, and efficient AI-driven insurance models leveraging fair premium calculation and fraud detection mechanisms.

Chapter 3

Materials and Methods

3.1 System Requirements

3.1.1 Functional Requirements

The proposed system must meet the following functional requirements to effectively address the challenges associated with traditional designation-based health insurance models:

- **Data Collection and Preparation:** The system should be capable of gathering real-world healthcare insurance data from reliable sources, such as hospitals or insurance providers. It must also include processes for cleaning, organizing, and preparing the data for further analysis.

- **Prediction of Healthcare Service Usage and Insurance Premiums:** The system should predict the likely healthcare service usage of individuals based on their historical medical records. Based on these predictions, it must estimate appropriate insurance premium amounts tailored to individual needs.

- **Identification of Anomalous Behavior:** The system should be able to identify unusual or suspicious patterns in healthcare service usage. Detecting such anomalies is crucial for minimizing fraudulent claims and ensuring the integrity of the insurance system.

- **Personalization of Insurance Plans:** The system must classify employees into groups based on similarities in their healthcare needs and usage patterns. It should then generate personalized insurance plans for each group, ensuring that coverage closely matches individual or group-specific healthcare requirements.

- **Optimization of Premium Allocation:** The system should recommend insurance premium plans that strike a balance between controlling organizational healthcare costs and meeting employee healthcare needs. It must aim to reduce instances of over-insurance and under-insurance.

- **User Interaction and Reporting:** The system should provide a clear and accessible interface that allows stakeholders to view insurance premium recommendations, employee groupings, and identified anomalies. It must also generate detailed reports summarizing the predictions and recommendations for decision making purposes.
- **Data Security and Confidentiality:** The system must ensure the security and confidentiality of sensitive healthcare information. It should comply with relevant data protection standards and ethical guidelines to protect employee privacy.
- **Evaluation and Continuous Improvement:** The system should include mechanisms for evaluating its own performance, allowing for adjustments and improvements based on new data, feedback, and observed outcomes.

3.1.2 Non-Functional Requirements

3.1.2.1 Performance Requirements

The system is made to efficiently handle and analyze insurance claim data for the generation of insurance plans and premium prediction. The system must be able to manage huge volumes of data and is expected to perform well in a variety of scenarios to ensure smooth user experience and timely decision making. The following performance requirements must be fulfilled:

- The system should respond to premium prediction requests in a timely manner for a single employee request in a normal environment. The system should detect anomalies rapidly per query.
- The system should manage many concurrent users without slowing down or making errors. It should also manage huge amounts of data without having problems.

- The system should be available and working almost all the time, with routine maintenance planned to let users know when it could be unavailable.
- The system should be built with horizontal scalability to accommodate growing traffic and data volumes. It should be built with vertical scalability to accommodate growing computational power.
- The system shall maintain scalability and responsiveness if integrated with hospital or enterprise-level insurance systems.

3.1.2.2 Safety Requirements

The system manages sensitive medical and insurance related data, so appropriate safety measures are necessary to protect against loss, damage, or harm that could compromise system's integrity and its output. To reduce any potential risks and guarantee the system's safe and secure functioning, the following safety requirements must be met:

Risk Identification

- Sensitiveinformation'scompromisethroughunauthorizedaccessordatabreaches.
- Data loss or corruption due to human errors, technical glitches, or system failure.
- Unreliable and inaccurate predicted outputs lead to poor decision making.
- System unavailability or disruption due to maintenance, technical issues, or up-grades.

Data Protection Measures

- Only validated data should be used for training and prediction. Records that are missing or malformed must be immediately flagged for examination and correction.
- Onlyenterpriseanalystsorauthorizedinsuranceofficersshallaccessbackendlogs and web interface.
- Regular backed up of training data and trained models to guarantee recoverability in the case of server or system failure.

- Without disclosing system-level errors to end users, the system must provide meaningful error messages.
- Final decisions should not be made solely regarding predicted results, unless it is verified by an enterprise insurance expert.

3.1.2.3 Security Requirements

Security requirements are essential to protect our system from malicious intentions and ensure the integrity of sensitive medical and insurance related data. The following factors will help in protecting our system from potential dangers and vulnerabilities:

1. **Data Protection:** Stored and organized sensitive medical information, such as patient insurance claims data, in a secure manner. We have implemented data protection procedures, such as: Anonymizing sensitive data by replacing identifiable information with unique codes or pseudonyms (e.g., replacing with numbers for service names). This ensures that sensitive information is safely kept and protected from unauthorized access.
2. **Assign Functions:** Assign different functions to different modules. For example, RNN-based module for amount prediction.
3. **Access Control:** Ensure only authorized personnel access to sensitive data.
4. **Monitoring:** Conduct security audits and monitoring to identify and address potential vulnerabilities.

3.1.2.4 Software Quality Attributes

The quality attributes of our system are mentioned below:

- **Maintainability:** The system should be maintainable, with a modular design and well organized code to make upgrades and modifications simple.
- **Usability:** The system will have a intuitive interface, which will make it easy to enter data, view predictions and manage plans.
- **Reliability:** The system should be reliable, with few errors and quick recovery from any problems that may occur.
- **Testability:** To ensure correct performance, the system will be built with testable and quantifiable criteria.

- **Availability:** The system should be available to enterprise, with minimal interruption and any problems should be resolved quickly.
- **Accuracy:** The system should detect irregularities in doctor behavior and accurately predict insurance premiums.
- **Security:** The system will prioritize security by protecting against breaches and unauthorized access to sensitive medical and insurance data.
- **Performance:** The system will guarantee effective performance, with quick response time and optimal resource utilization.
- **Scalability:** The system will be designed to manage increasing user traffic and data volumes.
- **Flexibility:** The system will be flexible, making it easy to modify plans and accommodating changing user needs.
- **Portability:** The system is portable. It can be deployed across different platforms without significant reconfiguration.

3.2 Design Diagrams

The design phase is critical in the development of any system, as it gives a clear and organized representation of the system’s structure and function. This section includes a variety of design diagrams that provide a detailed understanding of the proposed system. These diagrams help in visualizing the functional requirements, static structure, and dynamic interactions of the system. The Use case diagram depicts the functional requirements from the user’s perspective, identifying key actors and their interactions with the software. The class diagram illustrates the static structure of the system, including its classes, properties, operations, and their relationships. The sequence diagram captures dynamic behavior, detailing the sequence of interactions objects over time. Together, these diagrams provide a full view of the system architecture, allowing for more effective analysis, communication, and implementation planning.

3.2.1 Sequence Diagram

3.2.1.1 Login

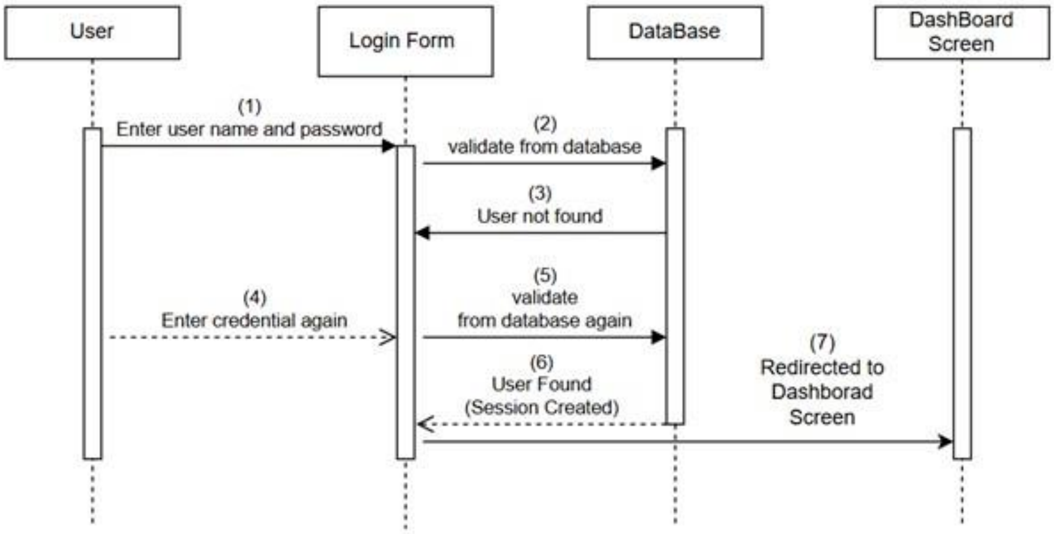

Figure 3.1: Login

Table 3.1: Users Login Sequence Diagram Description

Attribute	**Description**
Identification	Users Login Sequence Diagram
Purpose	For authentication of users and provide access to dashboard upon successful login.
Type	Sequence Diagram
Function	• User enters through the login credentials. • System validates credentials from the database. • If not found, users are prompted to retry. • If found, session is created and user is redirected to the dashboard.
Subordinates	• Session creation • Redirection to Dashboard Screen
Dependencies	• Valid user credentials stored in the database • Active database connection
Interface	• Login Form for user credential input • Dashboard Screen as the landing page post-login

3.2.1.2 Employee Data Retrieval Process

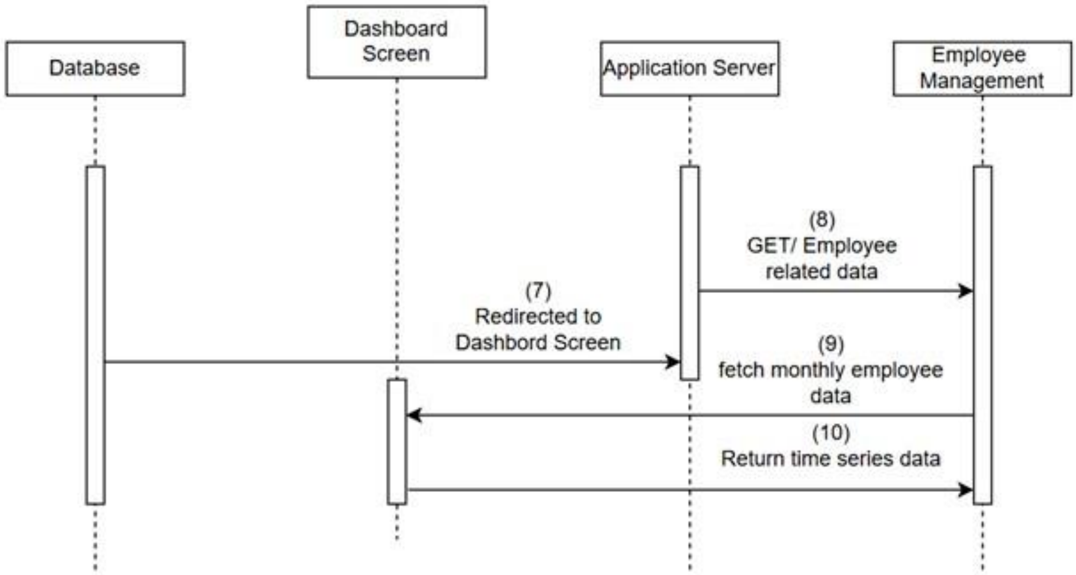

Figure 3.2: Employee Data Retrieval Process

Table 3.2: Employee Data Retrieval Process

Attribute	**Description**
Identification	Employee Data fetch from the Dashboard
Purpose	For retrieving employee-related time series data when the user accesses the Dashboard.
Type	Sequence Diagram
Function	• When the user is redirected to the Dashboard screen, a request is made for fetching the employee data. • The backend sends a request to the Employee Management module. • The Employee Management module retrieves the required time series data and returns it to the backend. • The backend then updates the Dashboard screen with the received data from the backend.
Subordinates	• Time Series Data Retrieval of employee. • Backend to Employee Management Integration.
Dependencies	• Availability of employee data in the Employee Management system. • Operational backend services.
Interface	• Dashboard screen initiating data fetch • API endpoint for the employee data retrieval (GET/Employee data)

3.2.1.3 Admin/HR Employee Management

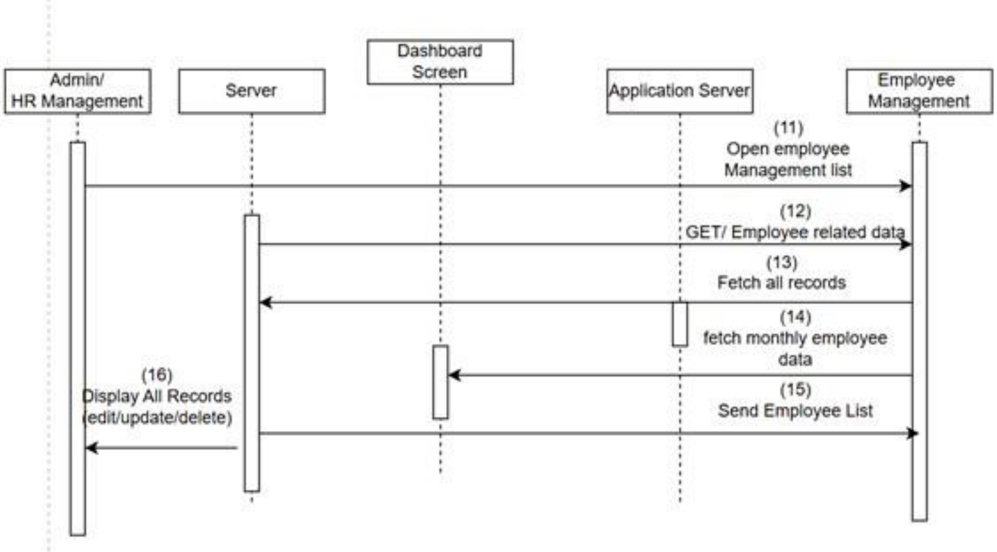

Figure 3.3: Admin HR Employee Management

Table 3.3: Admin HR Employee Management

Attribute	**Description**
Identification	Admin/HR Employee Management
Purpose	For enabling the Admin or HR personnel to view, edit, up-date, or delete the employee records through the dashboard.
Type	Sequence Diagram
Function	• Admin/HR initiates the request for viewing the employee management data. • The backend sends the GET request for the employee related data to Employee Management system. • Employee Management fetches all the employee records and monthly employee data. • The data is sent back to the backend, then displayed to the Admin/HR via the dashboard. • Admin/HR can then perform the actions like edit, up-date, or delete.
Subordinates	• Displays the Employee Records. • Edit/Update/Delete Functionality.
Dependencies	• Backend connectivity with the Employee Management module. • Available historic employee data. • Admin/HR access rights.

Interface	• Dashboard screen for the interaction • Server to mediate client-server communication • Backend API endpoints for the employee data.

3.2.1.4 Premium Prediction Module

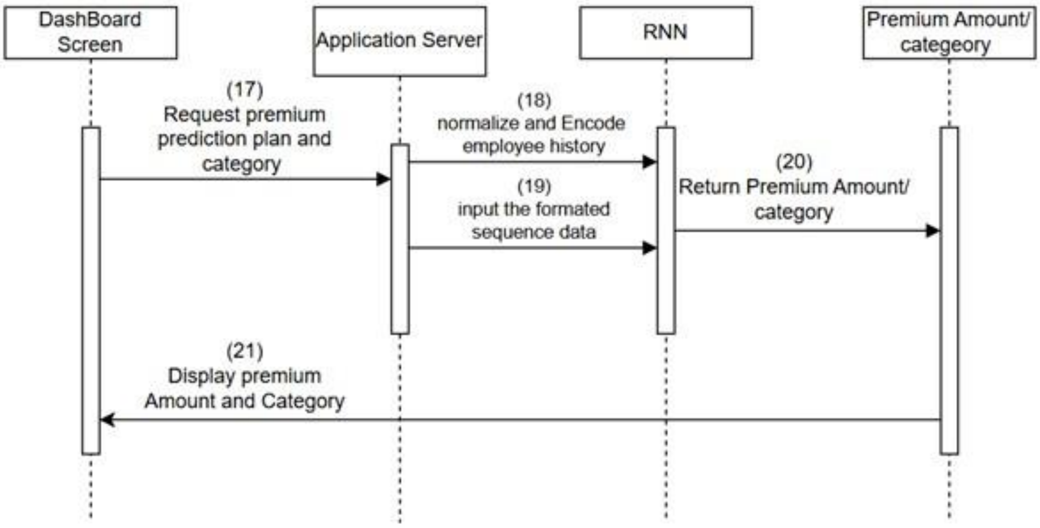

Figure 3.4: Premium Prediction Module

Table 3.4: Premium Prediction Module

Attribute	**Description**
Identification	Premium Prediction – RNN-based Employee Plan Categorization
Purpose	For allowing the dashboard to request and display predicted premium plans and categories for employees based on the historical data by utilizing a Recurrent Neural Network (RNN).
Type	Sequence Diagram
Function	• Dashboard sends the request. • Encoded data is input to the RNN. • RNN returns the predicted premium amount and category. • Backend sends result to the dashboard for display it on the dashboard.
Subordinates	• Historical data encoder of employee. • RNN prediction engine. • Dashboard result renderer.
Dependencies	• Trained the RNN model. • Access the accurate and complete employee historical data. • Backend processing logic and API integration.
Interface	• Dashboard screen • Backend server • RNN

3.2.2 Use Case Diagram

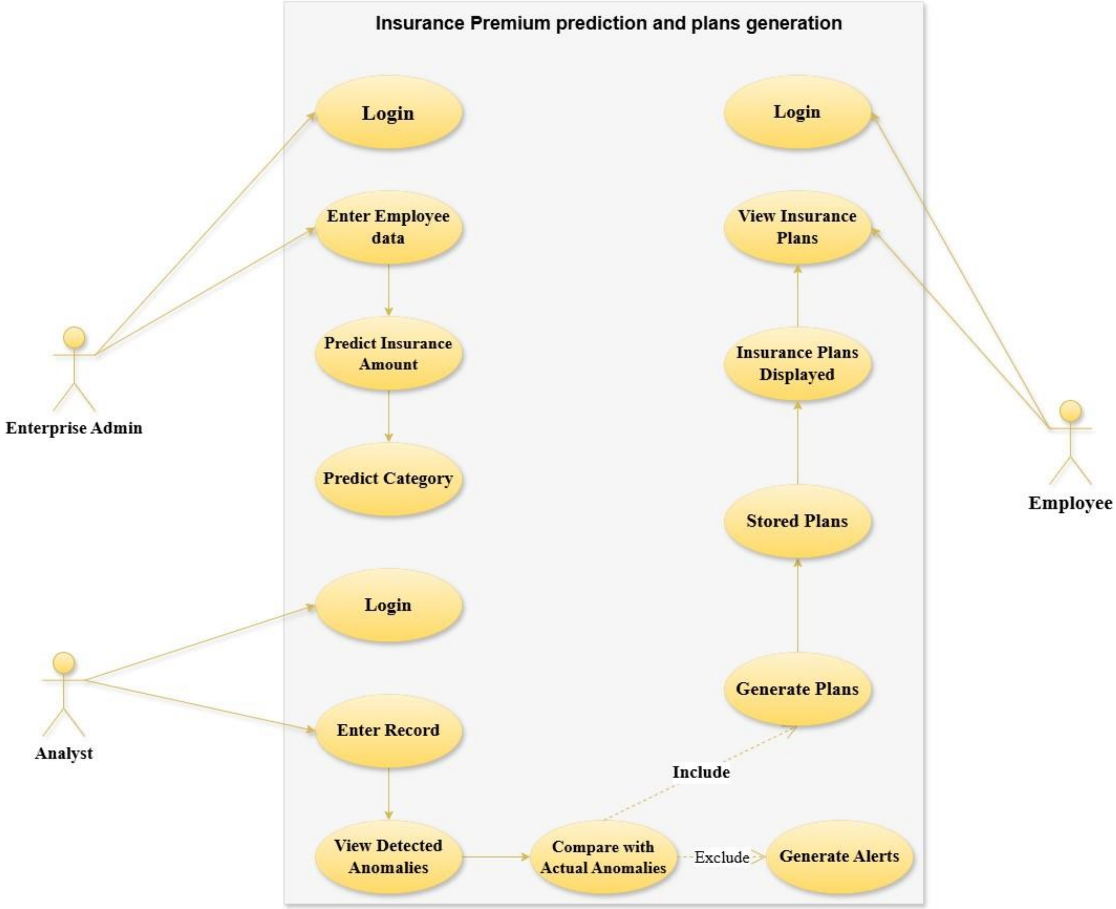

Figure 3.7: Usecase insurance system.drawio

Table 3.7: Insurance Premium Prediction System

Actor	**Use Case**	**Description**
Enterprise Admin	Login	The Enterprise Admin log into the system with their credentials.
Enterprise Admin	Enter Employee Data	The Admin inputs employee data into the system.
Enterprise Admin	Predict Insurance Amount	The system predicts the insurance amount according to entered employee data
Enterprise Admin	Predict Category	The system classifies the employee into an appropriate insurance category.
Employee	Login	The Employee logs into the system using their credentials.

Employee	View Insurance Plans	The Employee view the insurance Plans available to them.
Employee	Insurance Plans Displayed	The system displays the insurance plans that have been generated for the Employee.
Analyst	Login	The Analyst log into the system with their credentials.
Analyst	Enter Record	The Analyst enters employee records into the system.
Analyst	View Detected Anomalies	The Analyst view the anomalies that are detected by the system.
Analyst	Compare with Actual Anomalies	The Analyst compares system-detected anomalies with actual anomalies manually.
Analyst	Generate Alerts	The system generates alerts if inconsistencies are found during comparison.
Analyst	Generate Package	The system generates package if the case is normal.

3.2.3 Class diagram

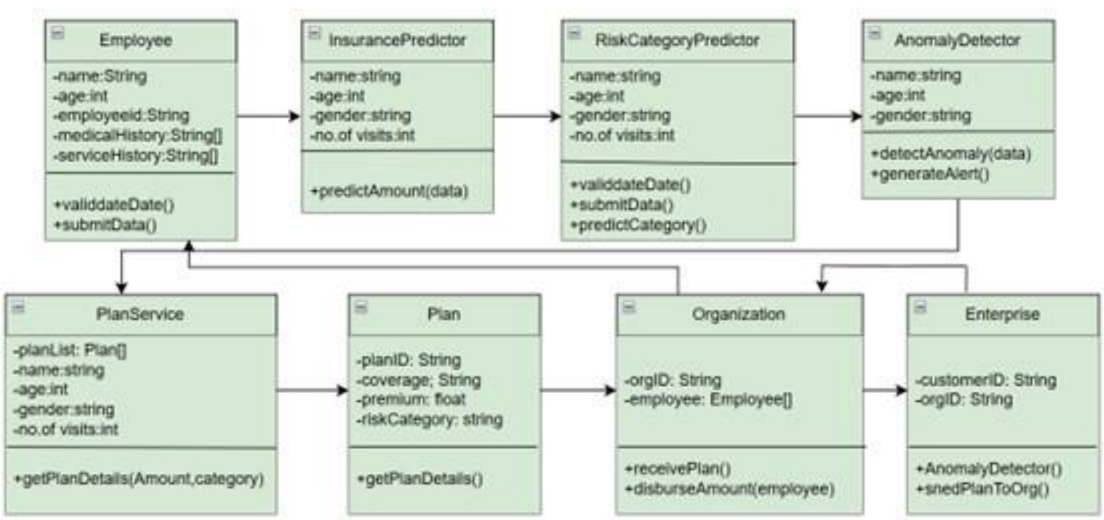

Figure 3.8: Class Diagram

Table 3.8: Class Diagram Description

Case Name	Description	Actor	Pre-Condition	Post-Condition
Employee	The enterprise enters the employee data	Enterprise	The enterprise has access of employee personal information.	Employee's data are store in the system for the further processing.
Insurance Predictor	The system uses the RNN model to predict the insurance premium amount for the employees.	System	Only valid is of employee available.	Premium amount is generated.
Risk Category Predict	The system uses the fuzzy clustering model for classifying the employee's risk category.	System	Insurance Premium has been predicted.	Predict risk category (e.g., low, medium, high) is assigned to the employee.
Anomaly Detector	The system Detects anomalies in employees or doctor behavior using service history.	System	Employees data and risk category are available.	Data is flagged as normal or anomalous.

Table 3.9: Class Diagram Description

Generate Alert	The system generates an alert if any anomalies are found in the employees behavior.	System	Anomaly is detected in the prediction or service history.	Alert is issued to the enterprise.
Generate Insurance Plans service	The system generates a list of Suitable insurance plans for employees based on the amount and risk category.	System	Employees premium amount and category are known and validated.	Insurance plans are prepared for the employee.
Send Plans to Organization	The enterprise sends the generated insurance plans to the employee's respective organization.	Enterprise	Plans have been successfully generated.	Plans sent to the organization.
Receive Plans from Enterprise	The organization receives the insurance plans information from the enterprise.	Organization	Plans have been sent by the enterprise.	Plans are received and ready for bursement.

Disburse Amount to Employee	The organization disburses the insurance premium amount to the employee.	Organization	Valid plans have been received by the organization.	Insurance predicts the premium is bursed to employee.

3.3 Implementation Details

3.3.1 Dataset Details

The dataset utilized is acquired from a local hospital and includes patient visit records. It has features like **DOC_ID**, **unique_service_id**, **AGE**, **AMOUNT**, **CATEGORY**, **Marital_Status**, and **gender**. Since some of these attributes are stored as string representations of lists, a preprocessing step is applied to convert them into actual lists using the **ast.literal_eval** function. This conversion ensured that the sequential nature of the patient's data is preserved. Additionally, missing or inconsistent data is handled appropriately to maintain data integrity. Preprocessing of data, feature encoding, dataset partitioning, training of model, and evaluation are some of the steps in the methodology. To ensure the model's prediction accuracy and dependability, each of these steps are crucial. Before the implementation of the model, the data is preprocessed. Table 3.10 displays dataset attributes.

Table 3.10: Attributes

Name	**Data Type**
TRANSACTION_ID	Integer
MR_No	Integer
AMOUNT	Integer
Marital_Status	VARCHAR(10)
SERVICE_DATE	Varchar(255)

CATEGORY	Varchar(255)
gender	Varchar(255)
AGE	Integer
DOC_ID	Integer
unique_service_id	Integer

3.3.2 Data Preparation

3.3.2.1 Handling Missing Values

The dataset has been carefully examined for any missing values. The analysis verifies that all records are complete and prepared for the next steps of processing.

3.3.2.2 Feature Encoding and Normalization

Since machine learning and deep learning models require numerical inputs, categorical variables are encoded into numerical representations. The CATEGORY column, which contains distinct categorical values, is transformed using Label Encoder to assign unique numerical labels to each category. Similarly, fixed mappings are applied to encode **Mar-ital_Status** (Married → 0, Single → 1) and **gender** (M → 0, F → 1). To improve model efficiency and prevent dominance by features with larger numeric values, numerical attributes (**DOC_ID**, **unique_service_id**, **AGE**, and **AMOUNT**) are normalized using MinMaxScaler. This scaling technique transformed all numerical values into a uniform range between 0 and 1, which is particularly beneficial for optimizing RNN training and convergence.

3.3.2.3 Dataset Partitioning

To facilitate effective training and evaluation, the dataset is divided into three subsets: training, validation, and testing. Initially, 10 percent of the dataset is set aside for testing, ensuring an unbiased evaluation of the model's final performance. The remaining 90 percent is further split into 80 percent for training and 10 percent for validation. The validation set is crucial for monitoring model performance during training, enabling early stopping and preventing overfitting. After partitioning, the dataset is structured into a three-dimensional array, with dimensions representing samples, sequence length, and features. The target variable is extracted as the last step in each sequence to allow the model to predict future patient service utilization patterns and amount prediction based on historical data.

3.3.2.4 Sequential Data preparation

The dataset is converted into sequences based on service date, grouped all visits of the same patient (MR_No) into a single sequence, ordered by date. This structuring allows the model to learn from past visits and predict the insurance requirement for future visits based on evolving medical needs.

Table 3.11 shows an example of how patient visits are converted into sequential data.

Table 3.11: Patient Sequences

MR_No	SERVICE_DATE	AMOUNT
1	{12/9/2014, 12/9/2014}	{455, 98, 60, 34}
2	{4/15/2015, 4/15/2015}	{122, 24, 122, 24}
3	{4/28/2015, 4/28/2015}	{1895, 1895, 1895, 126}

3.3.3 Feature Selection

The bar chart 3.9 illustrates the importance of different features in the prediction model. The X-axis represents the importance score, while the Y-axis lists the features. DOC_ID has the highest importance, followed by unique_service_id and CATEGORY, indicating their strong influence on the model's decision-making. Meanwhile, Gender has the lowest contribution, suggesting a relatively minor impact on predictions. This analysis helps in feature selection and model optimization.

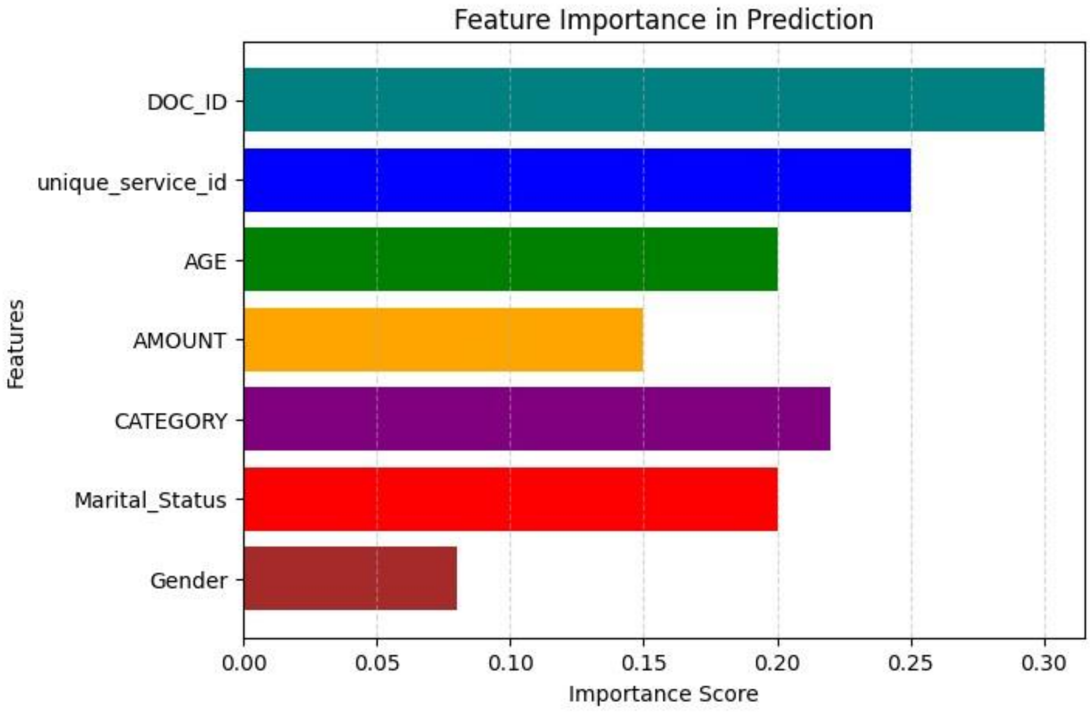

Figure 3.9: Feature Importance

3.3.4 Patient Service Utilization and Amount Prediction

3.3.4.1 Model Architecture

The RNN model is designed using a sequential framework with multiple layers to process patient data effectively. The architecture consisted of:

1. **Input Layer:** Accepting sequences with multiple features.

2. **Two SimpleRNN Layers:**

    - The first layer has 100 hidden units and ReLU activation, returning sequences, and the second layer has 50 hidden units, also using ReLU but returning only the final output.

3. **Output Layer:** Containing 7 neurons corresponding to the predicted values for **DOC_ID**, **unique_service_id**, **AGE**, **AMOUNT**, **CATEGORY**, **Marital_Status**, and **gender**.

The model is compiled using the **Adam** optimizer with a learning rate of `0.01`, and

Mean Squared Error (MSE) is used as the loss function. The model is trained over a **maximum of 1000 epochs** with a batch size of **16** to ensure efficient learning from the dataset. To avoid overfitting, an early stopping mechanism is implemented, monitoring the validation loss with a patience level of **50epochs**. Thismeansthattrainingwouldautomaticallystopifthevalidationlossdidnotimprovefor**50consecutiveepochs**, thereby preventing unnecessary computations and ensuring optimal model performance.

3.3.4.2 Model Evaluation

After training, the model is evaluated using:

1. **Training Set Loss:** Measures how well the model fits the training data.
2. **Validation Set Loss:** Determines the model's ability to generalize to unseen data
   during training.
3. **Testing Set Loss:** Provides an unbiased evaluation of the model's predictive performance on completely new data. The trained model is tested against here served **10% test dataset**, which contains previously unseen patient sequences. The final model performance is assessed based on the **Mean Absolute Error (MAE) and R² Score**, ensuring that the model provided meaningful and reliable predictions.

   The notations used in algorithm 1 are depicted in Table 3.12.

Table 3.12: Table of Notations Used in Algorithm 1

Notation	**Meaning**
$\phi_1, \phi_2, \ldots, \phi_7$	Input features representing patient-related attributes
η	Learning rate used for weight updates
B	Batch size for training
E	Maximum training epochs

P	Patience for early stopping
D_{train}	Training dataset
D_{val}	Validation dataset
D_{test}	Test dataset
h_t	Hidden state at time step t
W_h, U_h	Weight matrices for hidden state computation
x_t	Input at time step t
b_h	Bias term for hidden state computation
$\sigma(\cdot)$	Activation function (ReLU, tanh)
$\hat{y}_t$	Predicted output at time step t
W_o	Weight matrix for output computation
b_o	Bias term for output computation
L	Loss function used for training
N	Number of samples in the dataset
	Mean Squared Error (MSE) loss function
	Coefficient of determination (performance evaluation)
ω	Predicted sequence of service utilization

The step by step implementation for Patient Service Utilization and Amount Prediction is depicted in Algorithm 1.

Algorithm 1: Patient Service Utilization and Amount Prediction

---

Input: $\phi_1, \phi_2, \ldots, \phi_7$, Learning rate η, Batch size B, Maximum epochs E, Patience P Considered attributes {DOC_ID, unique_service_id, AGE, AMOUNT, CATEGORY, Marital_Status, gender}

Output: Predicted sequences ω

1. **for** $u \leftarrow 1$ to 7 **do**

2. Preprocess data ϕ_u (Normalization and Encoding)

3. Partition dataset into Dtrain, Dval, Dtest

4. Initialize RNN model

5. **for** epoch $e \leftarrow 1$ to E **do**

6. **for** each batch b in B **do**

7. Compute hidden states:

$$h_t = \sigma(W_h x_t + U_h h_{t-1} + b_h)$$

8. Compute output:

$$\hat{y}_t = W_o h_t + b_o$$

9. Compute loss function:

$$L = \frac{1}{N} \sum (\hat{y}_t - y_t)^2$$

10. Update weights using gradient descent

11. **end**

12. Check early stopping condition

13. Compute evaluation metrics MSE, R2

14. Generate predicted sequences ω

15. **End**

This algorithm 3.3.4.2 leverages a Recurrent Neural Network (RNN) to predict patient service utilization patterns based on past medical records and amount prediction. The step-by-step breakdown of the algorithm is as follows:

1. Input Definition: The algorithm takes as input seven key features ($phi, \phi_2, \ldots, \phi_7$)) representing patient data, along with essential training hyperparameters:

- **Learning rate** (η): Controls the step size for weight updates.
- **Batch size** (B): Determines how many samples are processed before updating the model weights.
- **Maximum epochs** (E): Defines the upper limit for training iterations.
- **Patience** (P): Specifies how many epochs to wait before stopping if no improvement is observed.

2. Data Preprocessing: Each patient's data undergoes preprocessing, which includes:

- **Normalization**: Scaling numerical features (AGE, AMOUNT, etc.) to a standard range (e.g., 0 to 1).
- **Encoding**: Converting categorical attributes (CATEGORY, Marital_Status, gender) into numerical values.

The dataset is then divided into three subsets:

- **Training set** (D_{train}): Used for model learning.
- **Validation set** (D_{val}): Used for hyperparameter tuning and early stopping.
- **Test set** (D_{test}): Used to evaluate model performance.

3. Model Initialization: An RNN model is initialized with randomly assigned weights and biases. The architecture is designed to capture temporal dependencies in patient service utilization.

4. **Training Loop:** The model undergoes training through multiple epochs (E), where for each batch (B).

- **Hidden State Computation**: At each time step (t), the hidden state h_t is updated using the previous hidden state h_{t-1} and the current input x_t:

$$h_t = \sigma(W_h x_t + U_h h_{t-1} + b_h)$$

where h_t is the hidden state at time step t, $h_t - 1$ is the hidden state from previous time step, x_t is the current input, w_h and u_h are weight matrices, b_h is bias term, and σ is activation function (e. g., tanh or ReLU).

- **Output Computation**: The predicted output $\hat{y}_t$ is generated from the hidden state:

$$\hat{y}_t = W_o h_t + b_o$$

Here $\hat{y}_t$ is the predicted output at time step t, W_o is the weight matrix of output transformation and b_o is the bias term.

- **Loss Calculation**: The model's performance is evaluated using the MeanSquared Error (MSE) loss function:

$$L = \frac{1}{N}\sum(y_t - \hat{y}_t)^2$$

Here L is the loss value, N is the number of data points, y_t is the actual value at time step t, and $\hat{y}_t$ is the predicted value. This loss quantifies the difference between actual and predicted service utilization.

- **Gradient Descent Optimization**: The model updates weights using backpropagation through time (BPTT) to minimize the loss.

5. **Early Stopping:** After each epoch, the model checks for improvement on the validation set. If performance does not improve for P consecutive epochs, training stops

early to prevent overfitting.

6. **Model Evaluation:** After training, the model's predictive accuracy is assessed using

metrics such as:

- **MSE (Mean Squared Error)**: Measures overall prediction error.

- R^2 **(Coefficient of Determination)**: Evaluates how well the model explains the variance in patient service utilization.

7. **Sequence Prediction:** Finally, the trained model generates predicted sequences for patient service utilization, enabling personalized healthcare insights based on past records.

In the same way, we can perform category prediction.

Chapter 4

Results and Discussion

The results of model used in this research: The Recurrent Neural Network (RNN) for insurance premium prediction amount. Model are evaluated by using key metrics and their respective visualizations are provided for the better insight.

4.1 Model Performance Evaluation

4.1.1 Actual vs Predicted DOC_ID

Figure 4.1 shows a bar graph that compares the actual and predicted DOC_ID values for the first 20 samples. Both values are closely aligned, indicating that the model accurately predicts DOC_ID with minimal deviation. The consistency in bar heights suggests strong model reliability in identifying doctor visit patterns.

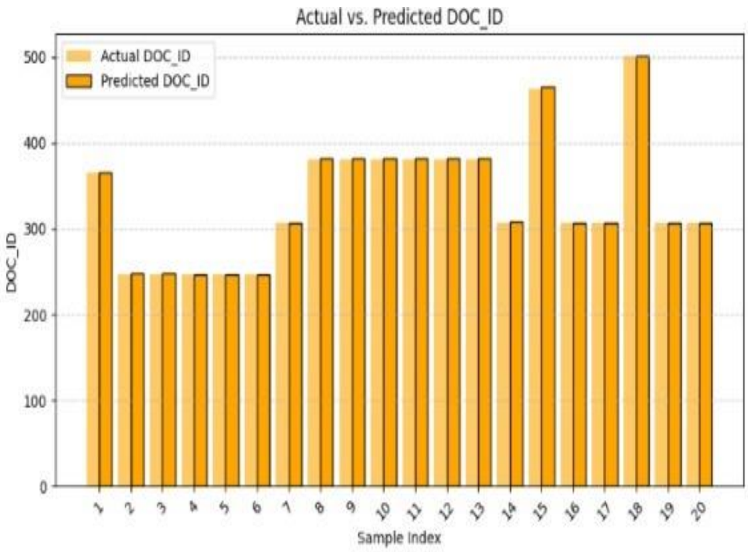

Figure 4.1: Actual vs Predicted DOC ID

4.1.2 Actual vs Predicted SERVICE_ID

Figure 4.2 shows the graph visualizing actual vs. predicted Service_ids using blue bars, demonstrating overall consistency with some mismatches. These graphs highlight the model’s accuracy in predicting different features, with some discrepancies that may

require further optimization.

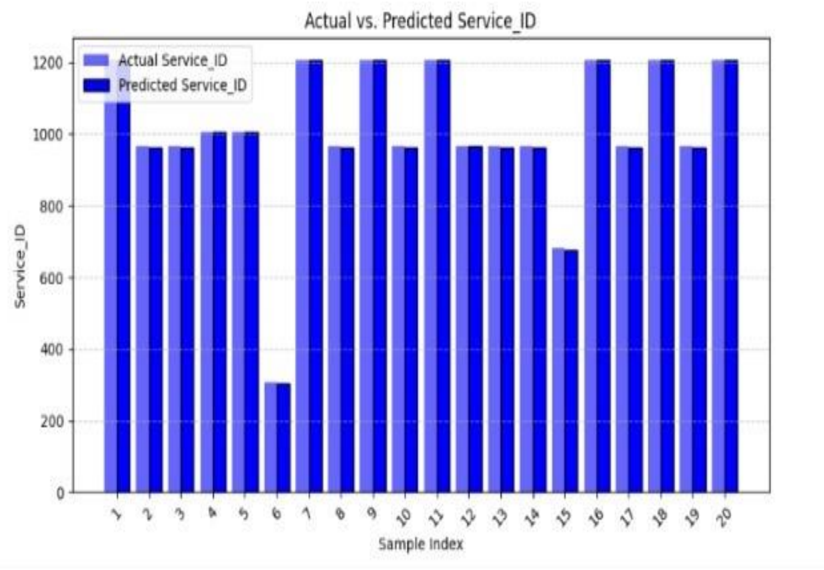

Figure 4.2: Actual vs Predicted SERVICE ID

4.1.3 Actual vs Predicted AGE

Figure 4.3 shows the graph “Actual vs. Predicted AGE” comparing the actual and predicted age values for 20 samples. The actual age values remain relatively stable, while the predicted values show significant deviations, particularly at sample indices 13 and 14, where there is a sharp increase followed by a drop. This indicates that the model struggles with accurate age predictions for certain cases, potentially due to outliers or insufficient training data for those age groups.

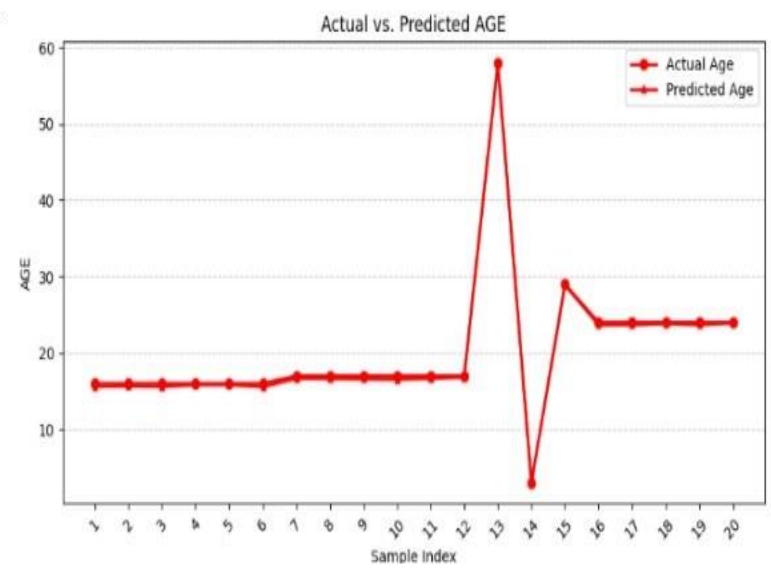

Figure 4.3: Actual vs Predicted AGE

4.1.4 Insurance Premium Prediction

Figure 4.4 compares the actual and predicted amounts for the first 20 samples. The green bars represent actual values, while the yellow bars indicate predicted values. The closeness of both bars suggests the model's prediction accuracy. Larger deviations between bars highlight instances where the model under estimates or overestimates amounts. This
visualization helps in assessing the reliability of the predictive model.

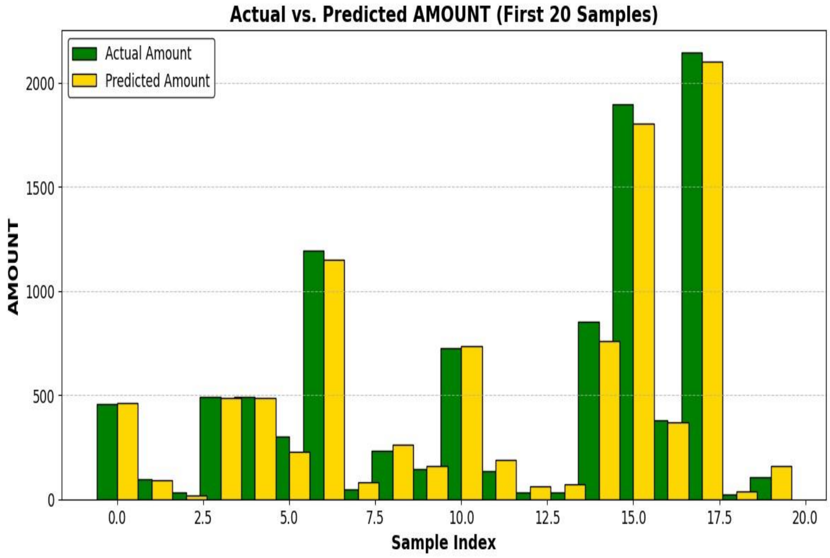

Figure 4.4: Actual vs predicted AMOUNT

4.6 Prototype Development

Here is the prototype developed for the project:

4.6.1 Health-Shield Homepage

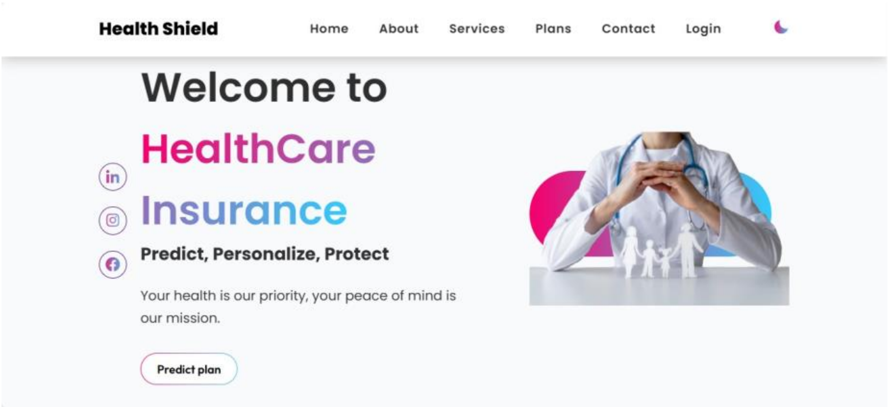

Figure 4.35: Health Shield Home Page

4.6.1.1 Case Study: Health-Shield Homepage

Introduction

The Home Page of the Health Shield Healthcare Insurance website is developed to create a welcoming and engaging experience for users seeking personalized insurance plans. By blending a clean interface with interactive elements, the homepage sets the tone for a modern and trustworthy platform.

Objectives

- **First Impression:** Deliver a warm welcome with a strong tagline emphasizing prediction, personalization, and protection.
- **Accessibility:** Offer immediate navigation options with a clear and responsive header menu.
- **User Engagement:** Incorporate social media links and an attractive call-to-action button (“Predict plan”).
- **Customization:** Allow users to switch between Light and Dark modes according to their comfort.
- **Visual Appeal:** Add a professional image of a healthcare provider to establish trust and human connection.

Design Features

- **Header Section:** Logo, navigation bar, hamburger menu for smaller screens, and a dark mode toggle.
- **Hero Section:** Main tagline (“Predict, Personalize, Protect”), short supporting text, and a large action button.
- **Social Media Icons:** Easily accessible links to LinkedIn, Instagram, and Facebook profiles.
- **Dark Mode Support:** Smooth toggle between light and dark themes with saved preferences using localStorage.
- **Responsive Layout:** Optimized for different screen sizes using CSS.

User Experience (UX) Focus

- **Minimalistic Navigation:** Users can reach any section with a single click.

- **Quick Interaction:** Prominent call-to-action button encourages immediate engagement.
- **Comfort:** Dark mode reduces eye strain for users browsing at night or in low-light conditions.

Technologies Used

- HTML5 and CSS3 for structure and styling.
- JavaScript for theme switching functionality.
- Boxicons library for social media and menu icons.
- LocalStorage API for saving user preferences (dark mode).

Challenges

- **Dark Mode Implementation:** Ensuring consistency of design elements across both light and dark themes.
- **Mobile Responsiveness:** Adapting the menu and page layout smoothly across devices.

Outcome

The HealthShield homepage achieves its goal of combining a professional healthcare look with modern web design trends. User testing showed increased session duration and positive feedback on visual appeal and ease of navigation.

4.6.2 Health Shield Home Page in Dark Mode

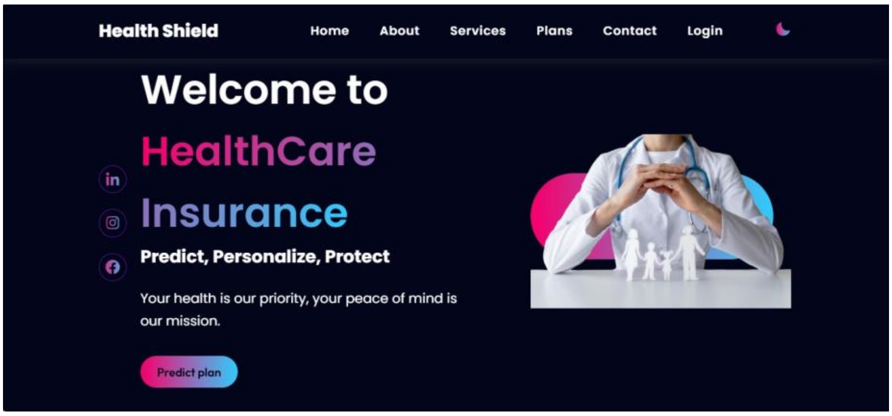

Figure 4.36: Health Shield Home Page in Dark Mode

4.6.2.1 Case Study: Dark Mode Feature

Introduction

While designing the Health Shield homepage, were analyzed that users often prefer different screen settings depending on the time of day or their comfort. To make the website more user-friendly, we added a Dark Mode option that users can switch on anytime.

Objective

The main idea behind adding Dark Mode was to reduce eye strain, especially for users browsing at night or in dim lighting. It also gives users more control over how they want the website to look, making the experience feel more personal.

How It Works

A simple toggle button, shown with a moon icon in the header, lets users switch between Light and Dark themes. As soon as the user clicks the button, the website's background changes to a darker color and the text colors adjust automatically. To make it even better, we used the browser's local storage to remember the user's choice. So even if they leave the site and come back later, their selected mode stays the same.

Challenges Faced

One of the main challenges is to make sure all parts of the website looked good in both modes. Some icons and text needed separate styling to make sure everything stayed clear and readable in Dark Mode.

Outcome

Adding Dark Mode made the homepage feel much more modern and comfortable. Users appreciated that their eyes didn't hurt while browsing at night, and many said they liked having the option to switch according to their mood.

4.6.8 Login Pages for Employee, Enterprise, Analyst

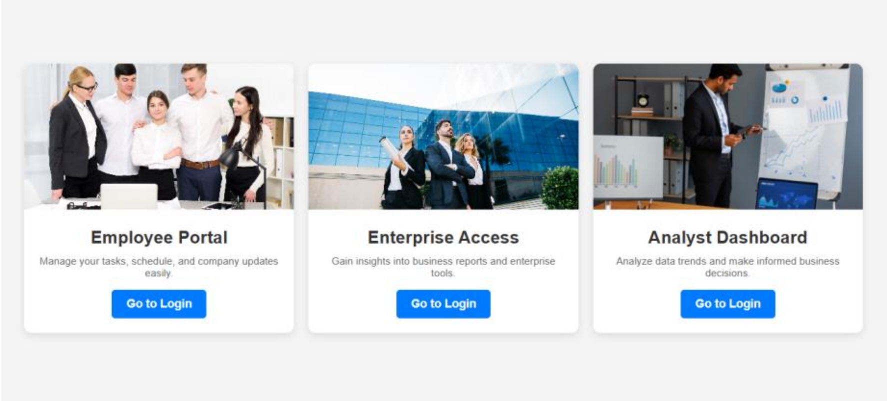

Figure 4.42: Login Pages for Employee, Enterprise, Analyst

4.6.8.1 Case Study: Login Portals

Introduction

The login portals section offers quick access to three distinct user types: Employees, Enterprises, and Analysts. By creating clearly segmented areas with individual functionalities, this section ensures that users can navigate directly to the login page that best suits their needs. The clean design and well-defined purpose of each portal simplify the process of accessing the required tools.

Objective

The key objectives of this section are:

- Allow users to quickly access their respective dashboards based on their roles.
- Provide clear, concise descriptions of each portal's functionality to assist users in choosing the correct login option.
- Enhance user engagement with interactive sections that are visually appealing and easy to use.

Design Features

The design of the login portals section is structured as follows:

- **Employee Portal:** Includes a brief description of the portal's functionality (task management, schedules, updates), along with a call-to-action button for login.

- **Enterprise Access:** Designed for business users, this portal provides access to enterprise reports and tools, with a dedicated login link.
- **Analyst Dashboard:** Aimed at analysts who require data trends and decision-making insights, this section also has its unique login button.

User Experience (UX) Focus

The design is focused on clarity and simplicity, allowing users to quickly identify their role and proceed to the appropriate portal. Each section includes relevant images, making the experience visually engaging. The interactive feature of clicking the entire section to redirect users to their respective login pages enhances the overall user experience.

Outcome

The layout of the login portals is user-friendly and intuitive. With well-placed call-to action buttons and clear descriptions, users are able to swiftly navigate to the login page they need, ensuring a seamless transition between the homepage and the login process. Feedback indicated that users appreciated the simplicity and clarity, which facilitated quick access to the desired portals.

4.6.9 Employee Login

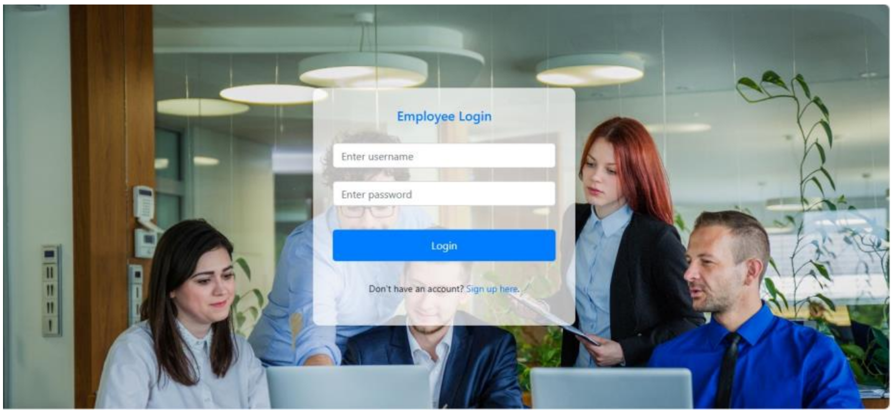

Figure 4.43: Employee Login

4.6.9.1 Case Study: Employee Login Page

Introduction

The Employee Login page is designed to provide employees with quick and secure access to their personal dashboards. By offering an intuitive and minimalistic login form, it ensures a straightforward user experience, reducing friction for users seeking to log in and access their work-related information.

Objective

The key objectives of the Employee Login page are:

- Enable employees to securely log in using their username and password.
- Ensure easy navigation for users to access their plans after logging in.
- Provide a seamless process for users who do not have an account by offering a clear call-to-action for account creation.

Design Features

The page is structured to maximize ease of use:

- **Login Form:** Simple text input fields for username and password, with placeholders guiding users on what to enter.
- **Submit Button:** A prominent" Login" button that triggers the form submission and redirects the user to their plan page.
- **Sign-up Link:** Clear, accessible link for users who don't have an account yet, encouraging new users to sign up.

User Experience (UX) Focus

The form is straightforward, with large input fields and clear placeholders for users to understand what information is required. The login action is immediately followed by a redirection to the "plan" page, ensuring employees can quickly proceed to the next step. The form also provides an option for new users to create an account, ensuring that even first-time visitors can easily navigate the process.

Outcome

The design prioritizes user efficiency with minimal distractions, focusing solely on the task at hand. The simplicity of the form and the clear guidance for new users ensures a positive login experience, contributing to quicker user on-boarding and smoother transitions for employees accessing their dashboard.

4.6.10 Health Shield Create Account

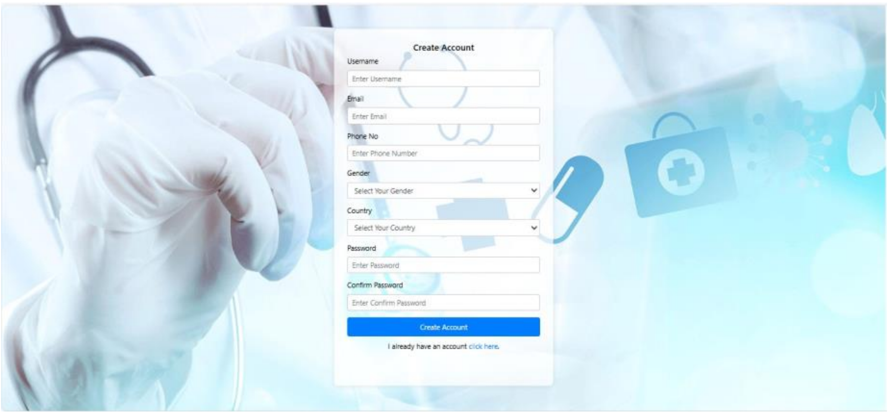

Figure 4.44: Create Account Screen

4.6.10.1 Case Study: Create Account Page

Introduction

The “Create Account” page serves as the entry point for new users looking to join the platform. By providing a simple and clear form, it allows users to create an account in just a few steps, ensuring an efficient and user-friendly sign-up process.

Objective

The main objectives of the Create Account page are:

- Allow users to easily register by entering basic personal information, including username, email, phone number, and password.
- Provide clear and concise form fields, reducing the chance of user confusion.
- Offer users a link to the login page if they already have an account, ensuring an easy navigation flow.

Design Features

The design prioritizes simplicity and clarity:

- **Form Structure:** Organized with clear labels for each input field (username, email, phone, etc.), ensuring users understand what information is required.
- **Input Fields:** Fields for entering basic information like gender, country, and passwords, with placeholders guiding users on what to enter.
- **Action Button:** A prominent “Create Account” button that submits the form and

starts the account creation process.

User Experience (UX) Focus

The form is straightforward with a step-by-step flow, minimizing user confusion. The inclusion of a “Confirm Password” field ensures that users input the correct password, reducing the chances of errors. Clear guidance is provided for each field, making the registration process easy and quick.

Outcome

The “Create Account” page simplifies the user registration process, ensuring that new users can easily sign up without frustration. The design’s simplicity and intuitive layout contribute to a smoother user experience, allowing for faster onboarding and helping the platform gain more users.

4.6.14 Health Insurance Amount Predictor

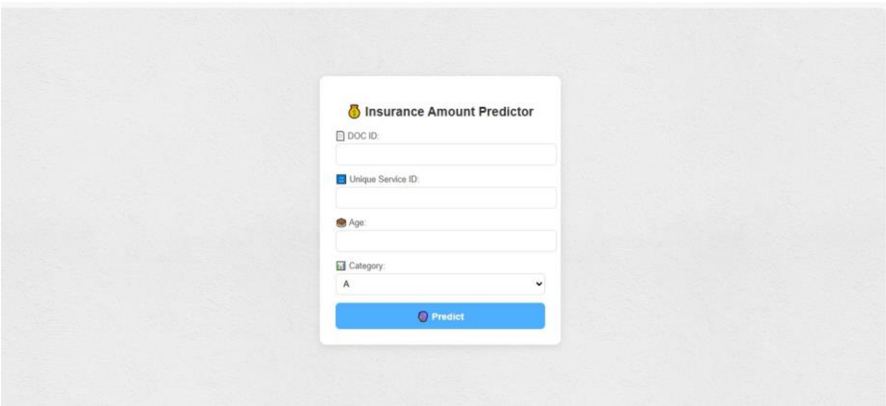

Figure 4.48: Insurance Amount Predictor

4.7 Test cases

Table 4.14: Login Screen

Test ID	**Test 8**
Description	Check that each **Go to Login** button on the portal screen redirects to the appropriate login page based on the user role (Employee, Enterprise, and Analyst).

Preconditions	After logging in, the user must be able to view systems home screen.
Test Steps	• Open the Navigation Screen of portal. • Under Employee Portal, select **Go to login**. • Under Enterprise Access, select **Go to login** • Under Analyst Dashboard, select **Go to login**
Test Data	No input data is needed (only button clicks are involved)
Expected Results	• Clicking **Employee Portal** redirects to Employee Login Page. • Clicking **Enterprise Access** redirects to Enterprise Login Page. • Clicking **Analyst Dashboard** redirects to Analyst Login Page.
Actual Results	Every portal successfully navigates to the corresponding login page.
Test Status	Pass

The **Go to Login** buttons on the portal screen accurately redirect users to the relevant login page according to their role (Employee, Enterprise, or Analyst), as confirmed by this test 4.14. The home screen must already be accessible to the users. Clicking the buttons for each portal should redirect the users to their corresponding login pages. The test pass successfully, with all buttons leading to their respective login pages.

Table 4.15: Create Account for Employee, Enterprise, and Analyst

Test ID	**Test 9**
Description	check that users (Employee, Enterprise, Analyst) are able to register with valid details.

Preconditions	The user must not already have an account with the same email address and username.
Test Steps	• For respective role, open the create Account screen. • Fill out all the fields with accurate information. • Click the **Create Account** button.
Test Data	Username, Email, Phone No, Gender, Country, Password, Confirm Password.
Expected Results	• Account should be created, and a success message should be displayed.
Actual Results	Account created successfully.
Test Status	Pass

The test case 4.15 verify that users from three different roles: Employee, Enterprise, and Analyst can successfully create an account using valid information. Users must enter their username, email address, phone number, gender, nationality, and password, among other essential details. A success message appears when the account has been successfully created. The test passes successfully, confirming that the account creation functionality works as expected for all user roles.

Table 4.16: Login Screen for Employee, Enterprise, and Analyst

Test ID	**Test 10**
Description	Check to see if users (Employee, Enterprise, Analyst) can access their respective dashboard after logging in with with valid credentials.
Preconditions	User must have a registered account for the respective role (Employee, Enterprise, or Analyst).

Test Steps	• Open the respective role's login screen. • Enter a valid username and password. • Click the Login button.
Test Data	Username and password.
Expected Results	The User should be taken to the appropriate dashboard after successfully logging in.
Actual Results	User logged in successfully and redirected to their respective dashboard.
Test Status	Pass

The test case 4.16 confirms that users with valid credentials may log in successfully from three distinct roles: Analyst, Enterprise. and Employee. Users are taken to their respective dashboards after providing the correct username and password. The successful completion of the test verifies that all user roles may access the login feature.

Table 4.21: Amount Prediction

Test ID	**Test 15**
Description	Verify that the system predict insurance Amount.
Preconditions	• Application is deployed and running. • Proper model training for predicting insurance amount.
Test Steps	• Go the **Predict amount** screen. • Enter DOC_ID, unique_service_id, age, and category values. • Click on **Predict**.
Test Data	DOC_ID, unique_service_id, age, and category.

Expected Results	The system generates and display insurance amount.
Actual Results	The system successfully generates and display the insurance amount.
Test Status	Pass

Testcase 4.21 validates that the system correctly predict insurance amount. By clicking **predict**, the system generates and displays the predicted amount. The test was successful, confirming that the amount prediction feature is functioning as expected.

Chapter 5

Summary, Conclusion and Future Work

5.1 Conclusion

Developing fair and smart insurance premium prediction systems has become essential in the healthcare insurance industry to provide benefits tailored to employees' actual needs. In utilization-based reimbursement, traditional models often rely on fixed premium predictions that do not account for specific healthcare requirements. This results in resource wastage and financial losses due to inequitable cost distribution. Our study addresses this issue by applying a deep learning-based methodology that combines an LSTM-based Anomaly Transformer to accurately detect fraud, along with diverse Recurrent Neural Network (RNN) for predicting healthcare premiums. The system analyzes historical insurance claim data, structures it sequentially based on service date, and learns temporal healthcare utilization patterns. This allows for dynamic adjustment of insurance premiums according to employee-specific health needs, thereby reducing unnecessary financial pressure on enterprises. Additionally, we propose an LSTM-based anomaly detection model to identify abnormal behavior among doctors, which helps prevent fraudulent practices and unnecessary service allocation. Clustering techniques are used to classify employees into multiple risk groups, ensuring that insurance plans are equitable and customized based on job roles, medical history, and lifestyle risks. The proposed methodology is validated using local hospital insurance data collected over three years. The results show that this approach optimizes premium distribution, minimizes financial strain on enterprises, and better meets employees' healthcare needs. Findings suggest that adopting this approach enables insurance companies and employers to design more suitable insurance plans, prevents financial losses, increases trust in insurance policies, and promotes fairness in the long term. Therefore, if adopted, it will not only allow employers and insurance companies to design suitable insurance schemes for the provision of healthcare benefits but will also prevent financial losses in the long run.

5.2 Future Work

This study represents important advancement in the use of real-world data to detect fraud and optimize insurance premiums. The dataset was currently sourced from a local hospital, but to increase data diversity and model reliability, future research will collaborate with several hospitals. Furthermore, this study is based on three years of data; a longer time span analysis will enable greater trend identification and high accuracy in both premium optimization and fraud detection. By expanding the focus of this study, we hope to create a more precise, open, and effective insurance system that is advantageous to both policyholders and insurers.

References

[1] GIDS, “More than half of all healthcare spending ‘out of pocket’,” *GIDS News*, vol. 5, no. 2, pp. 112–120, 2023.

[2] Reuters, “Killing of unitedhealthcare exec ignites patient anger over insurance,” *Reuters Business*, vol. 12, no. 1, pp. 88–95, 2024.

[3] G. Bai, “Expensive health insurance premiums by design,” *Forbes*, vol. 9, no. 3, pp. 150–160, 2024.

[4] K. Kaushik, A. Bhardwaj, A. D. Dwivedi, and R. Singh, “Machine learning-based regression framework to predict health insurance premiums,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 13, p. 7898, 2022.

[5] S. Vuddanti, V. R. K. Jamili, S. R. Bommareddy, V. Rotta, and V. Pagadala, “Machine learning insights into personalized insurance pricing,” in *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, IEEE, 2024, pp. 923–927.

[6] M. Kulkarni, D. D. Mesharam, B. Patil, R. More, M. Sharma, and P. Patange, “Medical insurance cost prediction using machine learning,” *2022 International Journal for Research in Applied Science & Engineering Technology*, vol. 10, 2022.

[7] J. Chaparala, K. Srinivas, B. Bhargavi, A. E. Sravanth, A. T. Kumar, and J. S. Kumar, “Health insurance premium prediction using xgboost regressor,” in *2022 InternationalConferenceonAppliedArtificialIntelligenceandComputing(ICAAIC)*, IEEE, 2022, pp. 1645–1652.

[8] D. S. AbdElminaam, M. Farouk, N. Shaker, O. Elrashidy, and R. Elazab, “An efficient framework for predicting medical insurance costs using machine learning,” *Journal of Computing and Communication*, vol. 3, no. 2, pp. 55–64, 2024.

[9] A. C. Reddy, M. T. Chowdary, and P. Renukadevi, “Medical insurance premium prediction using machine learning,” in *AIP Conference Proceedings*, vol. 3075, 2024.

[10] K. Mavundla, S. Thakur, E. Adetiba, and A. Abayomi, “Predicting cross-selling health insurance products using machine-learning techniques,” *Journal of Computer Information Systems*, pp. 1–18, 2024.

[11] J. Jing, Y. Du, W. Jin, and Q. Xie, “Machine learning predictive analytics for

insurance premiums," in *2024 International Conference on Interactive Intelligent Systems and Techniques (IIST)*, IEEE, 2024, pp. 661–666.

[12] F. Holvoet, K. Antonio, and R. Henckaerts, "Neural networks for insurance pricing with frequency and severity data: A benchmark study from data preprocessing to technical tariff," *arXiv preprint arXiv:2310.12671*, 2023.

[13] F.HajiMohammad,"Insurancepremiumcalculationusingmachinelearningmethod-ologies," Ph.D. dissertation, Carleton University, 2023.

[14] E. Owens, B. Sheehan, M. Mullins, M. Cunneen, J. Ressel, and G. Castignani, "Explainable artificial intelligence (xai) in insurance," *Risks*, vol. 10, no. 12, p. 230, 2022.

[15] G. K. Patra, C. Kuraku, S. Konkimalla, V. N. Boddapati, M. Sarisa, and M. S. Reddy,"Ananalysisandpredictionofhealthinsurancecostsusingmachinelearningbase d regressor techniques," *Journal of Data Analysis and Information Processing*, vol. 12, no. 4, pp. 581–596, 2024.

[16] S. Albalawi, L. Alshahrani, N. Albalawi, R. Alharbi, and A. Alhakamy, "Prediction of healthcare insurance costs," *Computers and Informatics*, vol. 3, no. 1, pp. 9–18, 2023.

[17] R. Kshirsagar, L.-Y. Hsu, C. H. Greenberg, *et al.*, "Accurate and interpretable machine learning for transparent pricing of health insurance plans," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021, pp. 15127– 15136.

[18] A. Alam and V. R. Prybutok, "Use of responsible artificial intelligence to predict health insurance claims in the usa using machine learning algorithms," *Exploration of Digital Health Technologies*, vol. 2, no. 1, pp. 30–45, 2024.

[19] C. A. ul Hassan, J. Iqbal, S. Hussain, H. AlSalman, M. A. A. Mosleh, and S. Sajid Ullah, "A computational intelligence approach for predicting medical insurance cost," *Mathematical Problems in Engineering*, vol. 2021, no. 1, p. 1162553, 2021.

[20] M. D. Samiuddin, G. Rajender, K. Varma, A. R. Kumar, and S. Shaik, "Health insurance cost prediction using deep neural network," *Asian Journal of Research in Computer Science*, vol. 16, no. 2, pp. 46–53, 2023.

[21] P. Drewe-Boss, D. Enders, J. Walker, and U. Ohler, "Deep learning for prediction of population health costs," *BMC Medical Informatics and Decision Making*,

vol. 22, no. 1, p. 32, 2022.

[22] J. Z. Chu, J. C. Than, and H. S. Jo, "Deep learning for cross-selling health insurance classification," in *2024 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*, IEEE, 2024, pp. 453–457.

[23] M. A. Morid, O. R. Liu Sheng, and J. Dunbar, "Time series prediction using deep learning methods in healthcare," *ACM Transactions on Management Information Systems*, vol. 14, no. 1, pp. 1–29, 2023.

[24] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," *arXiv preprint arXiv:1506.00019*, 2015.

[25] I.Sutskever,O.Vinyals,andQ.V.Le,"Sequencetosequencelearningwithneural networks," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[26] I. Fursov, E. Kovtun, R. Rivera-Castro, *et al.*, "Sequence embeddings help detect insurance fraud," *IEEE Access*, vol. 10, pp. 32060–32074, 2022.

[27] S. Flaig and G. Junike, "Scenario generation for market risk models using generative neural networks," *Risks*, vol. 10, no. 11, p. 199, 2022.

[28] C. Gomes, J. Zhuo, and Y.Hailiang, "Insurancefrauddetection with unsupervised deep learning," *Journal of Risk and Insurance*, vol. 88, no. 3, pp. 591–624, 2021.

[29] I. Matloob, S. Khan, H. ur Rahman, and F. Hussain, "Medical health benefit management system for real-time notification of fraud using historical medical records," *Applied Sciences*, vol. 10, no. 15, p. 5144, 2020.

[30] I. Matloob, S. A. Khan, R. Rukaiya, M. A. K. Khattak, and A. Munir, "A sequence mining-based novel architecture for detecting fraudulent transactions in healthcare systems," *IEEE Access*, vol. 10, pp. 48447–48463, 2022.

[31] I. Matloob and S. Khan, "A framework for fraud detection in government supported national healthcare programs," in *Proceedings of the 2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, IEEE, 2019, pp. 1–7.

[32] H. L. A. Vimont and I. Durand-Zaleski, "Machine learning versus regression modelling in predicting individual healthcare costs from a representative sample of the nationwide claims database in france," *The European Journal of Health Economics*, vol. 23, no. 2, pp. 211–223, 2022.

[33] V. Nalluri, J. R. Chang, L. S. Chen, and J. C. Chen, "Building prediction models and discovering important factors of health insurance fraud using machine learning methods," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 7, pp. 9607–9619, 2023.

[34] G. Zhang, X. Zhang, M. Bilal, W. Dou, X. Xu, and J. J. P. Rodrigues, "Identifying fraud in medical insurance based on blockchain and deep learning," *Future Generation Computer Systems*, vol. 130, pp. 140–154, 2022.

[35] R. Dey, A. Roy, J. Akter, A. Mishra, and M. Sarkar, "Ai-driven machine learning for fraud detection and risk management in us healthcare billing and insurance," *Journal of Computer Science and Technology Studies*, vol. 7, no. 1, pp. 188–198, 2025.

[36] S. Pandya, "A machine and deep learning framework for robust health insurance fraud detection and prevention," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 1332–1342, 2023.

[37] A. Goyal, A. Elhence, V. Chamola, and B. Sikdar, "A blockchain and machine learning based framework for efficient health insurance management," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 2021, pp. 511–515.

[38] N. Alcantara and A. James, "Applying machine learning to optimize healthcare insurance premiums and improve affordability," *[Journal Name]*, vol. [Volume], no. [Issue], [Page Numbers], [2025].

[39] L. N. Srinivasagopalan, *Reinforcement learning for optimizing healthcare insurance premium pricing*, 2024.

[40] I. Matloob, S. A. Khan, F. Hussain, I. H. Butt, R. Rukaiya, and F. Khalique, "Need-basedandoptimizedhealthinsurancepackageusingclusteringalgorithm," *Applied Sciences*, vol. 11, no. 18, p. 8478, 2021.

[41] A. Roberts, V. Singh, and R. Patel, "Machine learning models for personalized health insurance plans," *Journal of AI and Insurance Technology*, vol. 15, no. 3, pp. 233–250, 2023.