PreMedix-Medical-Insurance-Prediction-System

REPORT OF PROJECT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ARTIFICIAL INTELLIGENCE

Batch

(2022-2026)

SUBMITTED BY

Rosy

12200724

SUPERVISED BY

Dr. Ridhi Kapoor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DAV UNIVERSITY JALANDHAR-PUNJAB144012

JALANDHAR -144001, PUNJAB

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my project guide, **Dr. Ridhi Kapoor**, for their valuable guidance, constructive feedback, and continuous support throughout the duration of this project. I am also thankful to **Dr. Rahul Hans** and the Department of **Computer Science and Artificial Intelligence** for providing the necessary resources and an encouraging academic environment.

My appreciation extends to all faculty members and classmates whose suggestions and cooperation helped strengthen the quality of this work. Lastly, I am deeply grateful to my family for their constant encouragement and support, which motivated me to complete this project successfully.

DECLARATION

I, **Rosy**, hereby declare that the work which is being presented in this project titled "**PreMedix-Medical-Insurance-Prediction-System"** by me, in partial fulfilment of the requirements for the award of Bachelor of Technology (B. Tech) Degree in "Computer Science and Artificial Intelligence" is an authentic record of my own work carried out under the guidance of Dr. **Ridhi Kapoor.**

To the best of my knowledge, the matter embodied in this report has not been submitted to any other University/ Institute for the award of any degree or diploma.

Rosy
12200724

ABSTRACT

The primary objective of the project was to leverage machine learning techniques to accurately estimate an individual's annual medical insurance premium based on key demographic and health indicators.

The methodology involved comprehensive data preprocessing, exploratory data analysis, and the selection of an advanced regression model, specifically XGBoost (Extreme Gradient Boosting), due to its superior performance in handling complex, non-linear relationships within the dataset. The system was implemented using Python, with Streamlit providing the interactive front-end application for data entry, customer viewing, and real-time premium prediction. The final model demonstrated robust predictive accuracy, significantly improving the efficiency and consistency of premium estimation compared to traditional manual assessment methods. The analysis also provided valuable insights into the primary drivers of insurance costs, such as smoking status and BMI. This project successfully achieved its objectives by delivering a functional, data-driven tool ready for deployment in a contemporary insurance advisory setting.

Table of Content

1. Introduction

The rapid evolution of digital healthcare systems has created the need for intelligent, efficient, and user-friendly applications that can manage patient information and support decision-making. Medical insurance companies, hospitals, and health-tech startups increasingly rely on automated systems to store customer records, analyze health-related variables, and predict insurance premiums. The **PreMedix Insurance System** developed in this project is designed to address these requirements by providing a streamlined platform for managing customer health data and performing AI-driven insurance premium predictions. Built using Python and the Streamlit framework, the system integrates data collection, storage, machine learning inference, and dynamic visualization, creating a comprehensive solution for medical insurance analytics.

At its core, the system offers three essential functionalities: **adding customer data**, **predicting insurance premiums**, and **viewing stored customer records**. Each of these components is implemented through separate Streamlit modules, ensuring modularity, clarity, and ease of maintenance. The user interface is intuitive and professionally styled, supported by a custom theme defined in the configuration file

config

. The design emphasizes usability and visual consistency, providing a clean and modern look suitable for professional environments such as clinics, insurance offices, and corporate dashboards.

The first major module, **Add Customer Details**, allows administrators or system users to input a wide variety of customer attributes through an interactive form. This module, implemented in the _Add_Customer.py file

_Add_Customer

, collects demographic information such as name, age, gender, phone number, and region. It also gathers essential medical indicators such as BMI, blood pressure level, diabetes status, alcohol consumption, smoking habits, and exercise patterns. These variables are crucial for determining health risk and insurance pricing. The data is validated and securely stored in a CSV file (customers.csv), enabling persistent storage without requiring a complex database. This ensures easy accessibility, platform independence, and simplicity when handling

customer datasets.

The second core component is the **Predict Insurance Premium** module, implemented in the _Predict_Premium.py file

_Predict_Premium

. This part of the system integrates a pre-trained machine learning model (insurancemodelf.pkl) to calculate predicted insurance premiums based on a subset of customer attributes. The model uses critical health determinants such as age, BMI, number of children, and smoking status to estimate the premium. These variables have a strong statistical correlation with medical expenses and risk, making them suitable for predictive modeling.

In the prediction workflow, the user first enters a customer's phone number, which acts as a unique identifier. The system then retrieves the corresponding customer's stored information and displays it in a well-structured format. Once retrieved, the model processes the relevant numerical values and generates a premium prediction. The system also categorizes the customer's risk level into **Low**, **Medium**, or **High**, based on the predicted cost. This provides an immediate, actionable insight for insurance agents or healthcare administrators who rely on risk assessments to make informed decisions.

One of the unique features of this module is the ability to **save the predicted premium back into the customer record**, ensuring the system maintains an updated and complete dataset. This feedback loop enhances the utility of the system by ensuring that each customer's file contains both raw input and AI-generated insights.

The third module, implemented in the _View_Customer.py file

View_Customer

, enables users to view all stored customer records in a tabular format. This functionality is essential for data analysis, auditing, and record management. Insurance companies and medical administrators often need quick access to customer data for reviews, compliance, or reporting. The system loads the customers.csv file and displays it using Streamlit's dataframe visualization, offering a neat and readable interface with scrollable and sortable tables.

All modules are seamlessly connected through a navigable interface defined in the main application file app.py

app

. This file sets up the Streamlit page structure, applies custom styling, and showcases the core features using visually appealing cards. The home page also acts as the central hub from which users can access the three main modules. Thoughtful UI design ensures that the system remains user-friendly even for non-technical users, reinforcing the practicality of the application in real-world medical insurance contexts.

The configuration file config.toml

config

contributes to the professional look and feel of the project by establishing a consistent color palette and typography across the application. This improves readability and enhances the overall user experience.

Beyond its practical functions, the PreMedix Insurance System also demonstrates key principles in modern software development: modular coding, data persistence, user-centric design, and integration of machine learning into interactive applications. The separation of features into dedicated pages improves modularity, while the usage of a CSV-based backend simplifies deployment and makes the application lightweight. This level of architecture is particularly suitable for small to medium-sized enterprises and startups that require powerful functionality without the overhead of complex database systems.

The inclusion of a machine learning model elevates the system from a simple data management tool to an intelligent decision-support application. Predicting medical insurance premiums traditionally required detailed actuarial analysis, domain expertise, and manual effort. By automating this process using a trained model, the PreMedix system brings speed, efficiency, and consistency to premium analysis. This not only reduces operational costs for insurance companies but also supports fairness and transparency by removing human bias from pricing decisions.

Overall, the PreMedix Insurance System provides a complete, end-to-end solution for customer information management and premium prediction. It showcases how modern technologies such as Python, machine learning, and Streamlit can be used to create functional, aesthetically appealing, and user-friendly applications. It serves as an excellent demonstration of integrating AI into practical service environments, offering substantial value to medical insurance providers and customers alike.

2. About the Project

The **PreMedix Insurance System** is an innovative AI-powered web application designed to modernize the way insurance companies and healthcare providers manage customer information and estimate medical insurance premiums. Built using **Streamlit**, **Python**, and a trained **machine learning model**, this project transforms traditional manual calculations into a smart, automated, and highly efficient digital process. Its clean interface, smooth navigation, and automated predictions make it a powerful tool for real-world insurance operations.

At its core, the project focuses on **three major functionalities**—collecting customer data, predicting insurance premiums, and viewing all stored records. By combining simplicity with intelligence, the system aims to reduce human effort, minimize errors, and provide instant insights based on customer health data.

Key Highlights of the Project

1. Smart Customer Data Collection

- Add detailed customer information through an interactive form
- Stores vital information such as:
    - Age, gender, region
    - BMI, exercise level, alcohol habits
    - Diabetes, smoking status, blood pressure
- Automatically saves records to a structured CSV database
- Ensures clean, reliable, and ready-to-use data

2. AI-Based Premium Predictio

- Uses a trained machine learning model to estimate premiums
- Identifies customers via phone number
- Predicts premium using key factors:
    - Age

    - BMI
    - Children
    - Smoking habit
- Calculates **risk level**: Low, Medium, or High
- Allows saving predictions back into customer records

3. Centralized Customer Database Viewer

    - Displays all stored customer records in a clean table
    - Helps administrators quickly check, verify, and analyze data
- Provides transparency for insurance teams

Why This Project Stands Out

The PreMedix system doesn't just store data—it transforms it into **actionable insights**. Insurance companies often face challenges like inaccurate manual calculations, scattered data, and long evaluation times. This project solves all these problems using:

- **Automation**
- **AI prediction models**
- **Clean and user-friendly UI**
- **Instant risk analysis**

Project Vision

The main intention behind this project is to create a **smart, accessible, and intelligent insurance management tool** that assists organizations in:

- Reducing manual workload
- Improving decision-making with real data
- Offering fair and accurate premium estimates
- Enhancing customer experience with quick services

3. Objectives

The primary objective of this report is to present the design, development, and implementation of the **PreMedix Insurance System**, an AI-driven platform for medical insurance analysis. The report aims to clearly document the technical approach, functional components, and practical relevance of the system. The specific objectives are as follows:

Main Objective

- **To develop an intelligent system** capable of predicting medical insurance premiums using machine learning algorithms.
- **To design a user-friendly interface** using Streamlit for smooth navigation and effortless data entry.
- **To automate the process of collecting customer health information**, ensuring accuracy, consistency, and secure storage.
- **To create a centralized customer database** that can be easily viewed, updated, and managed.
- **To integrate predictive analytics** into real-time customer evaluation for faster and more reliable decision-making.
- **To categorize customers into risk levels** (Low, Medium, High) based on predicted insurance premiums.
- **To reduce manual errors and effort** involved in traditional insurance premium calculations.
- **To demonstrate the use of AI in practical applications**, showing how machine learning can improve efficiency in the insurance domain.

Technical Objectives

- Implement data handling and storage using CSV files for simple yet effective database management.

- Utilize a trained **machine learning model** to predict insurance costs based on key health factors.
- Build separate modules for adding customers, predicting premiums, and viewing records to maintain modular architecture.
- Ensure seamless interaction between frontend forms, backend logic, and stored data.
- Maintain consistent visual design through custom themes for a professional UI experience.

Research Analysis Objectives

- To explore the key health parameters affecting insurance premiums such as age, BMI, smoking habit, and lifestyle factors.
- To analyze how machine learning enhances insurance accuracy compared to traditional estimation methods.
- To evaluate the efficiency, reliability, and accuracy of the developed system.

4. Steps to Achieve Objectives

To successfully achieve the objectives of the PreMedix Insurance System, a systematic and structured approach was followed. The entire development process was divided into logical stages to ensure clarity, accuracy, and smooth implementation. The following steps describe how each objective of the project was fulfilled:

1. Requirement Analysis & Planning

- Identified the problem: difficulty in manual insurance premium calculations.
- Studied key parameters affecting medical insurance costs (age, BMI, smoking, children, etc.).
- Defined core modules needed:
    - Add Customer
    - Predict Premium
    - View Customers
- Selected suitable technologies: Python, Streamlit, CSV storage, ML model.
- Designed the scope and boundaries of the system.

2. Data Collection & Preparation

- Collected historical insurance data for training the machine learning model.
- Cleaned the dataset by handling missing values, normalizing inputs, and converting categorical data.
- Selected relevant features for prediction (age, BMI, children, smoker).
- Split the dataset into training and testing sets to ensure model accuracy.

3. Model Training & Evaluation

- Trained a machine learning model using algorithms such as Linear Regression / Random Forest.
- Evaluated model performance with metrics like MAE and R^2 score.
- Tuned parameters to improve prediction accuracy.

- Saved the final trained model as **`insurancemodelf.pkl`** for integration.

4. Designing the Application Architecture

- Planned a modular structure with separate files:
  - `_Add_Customer.py`
  - `_Predict_Premium.py`
  - `_View_Customer.py`
  - `app.py`
- Chose CSV files (`customers.csv`) for lightweight data management.
- Applied custom UI themes using `config.toml` for a professional interface.

5. Developing the Add Customer Module

- Created input forms for all customer details (name, age, gender, BMI, lifestyle habits).
- Implemented validation checks to avoid missing or incorrect data.
- Designed a script to store data automatically into the CSV file.
- Ensured the module supports continuous data updating.

6. Developing the Premium Prediction Module

- Loaded the machine learning model (`insurancemodelf.pkl`).
- Implemented search-by-phone-number to fetch stored customer details.
- Automatically extracted ML-required inputs (age, BMI, children, smoker).
- Generated real-time premium predictions.
- Assigned risk levels (Low, Medium, High) based on output values.
- Added functionality to save predictions back to the customer record.

7. Building the Customer Viewing Module

- Loaded `customers.csv` into a clean table format.
- Enabled scrolling, filtering, and easy visualization.
- Ensured error handling when no data is available.

8. Designing the Homepage & Navigation

- Used `app.py` to create the main dashboard with feature cards.
- Linked all modules through the Streamlit sidebar for smooth navigation.
- Applied branding like "□ PreMedix System" for a polished look.

9. Testing the Complete System

- Conducted functional testing for all modules.
- Checked prediction accuracy and data consistency.
- Validated the UI flow and responsiveness.
- Ensured that new customer entries instantly reflect in the prediction and viewing modules.

10. Documentation & Final Report Preparation

- Documented the system workflow, architecture, and algorithms used.
- Prepared detailed explanations of objectives, methodology, and results.
- Ensured proper screenshots, code references, and explanations were included.

5.Codes and Implementation

The implementation of the **PreMedix Insurance System** is carried out using Python and the Streamlit framework, with additional integration of a machine learning model for predicting insurance premiums. The project follows a modular structure where each functionality is implemented in a separate file, ensuring maintainability, clarity, and scalability. The major components include:

- **Add Customer Module**
- **Premium Prediction Module**
- **View Customer Module**
- **Main Application Interface**
- **Theme and UI Configuration**

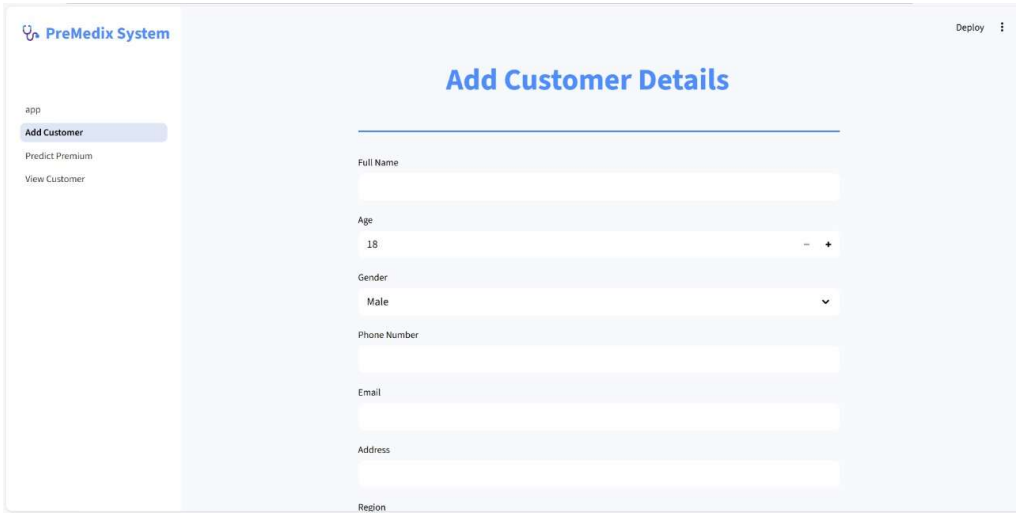
Each component plays a specific role in the system. The use of CSV files for storage and a `.pkl` file for the machine learning model ensures lightweight and efficient data handling.

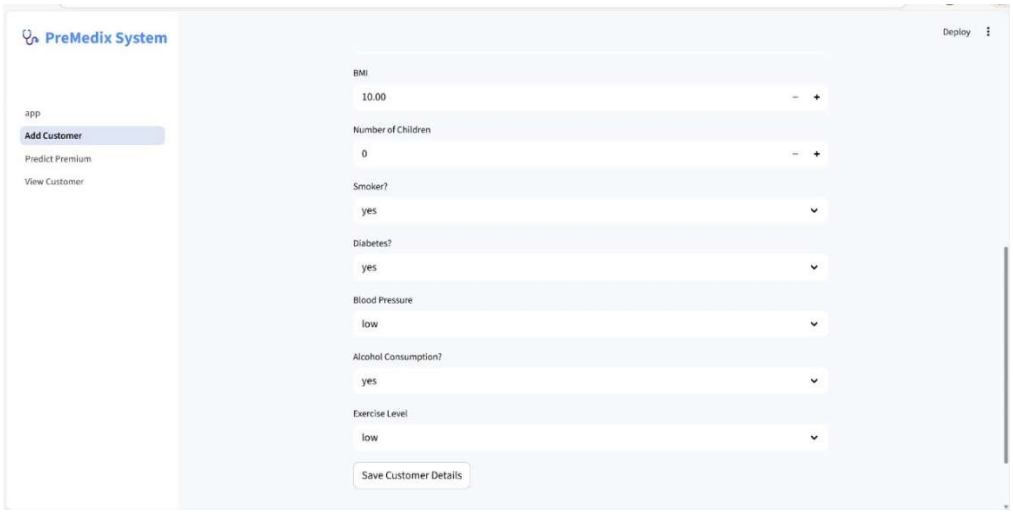
1. Add Customer Module (`_Add_Customer.py`)

This file implements the functionality for adding and storing customer details.
The core steps in the implementation include:

- Designing input fields using Streamlit widgets such as `text_input`, `selectbox`, and `number_input`.
- Collecting demographic and medical information including age, BMI, smoking, diabetes, blood pressure, and lifestyle.
- Checking whether the CSV database exists; if not, a new one is created with predefined columns.
- Appending new customer entries into **`customers.csv`** after validation.

- Displaying success messages after successful data submission.





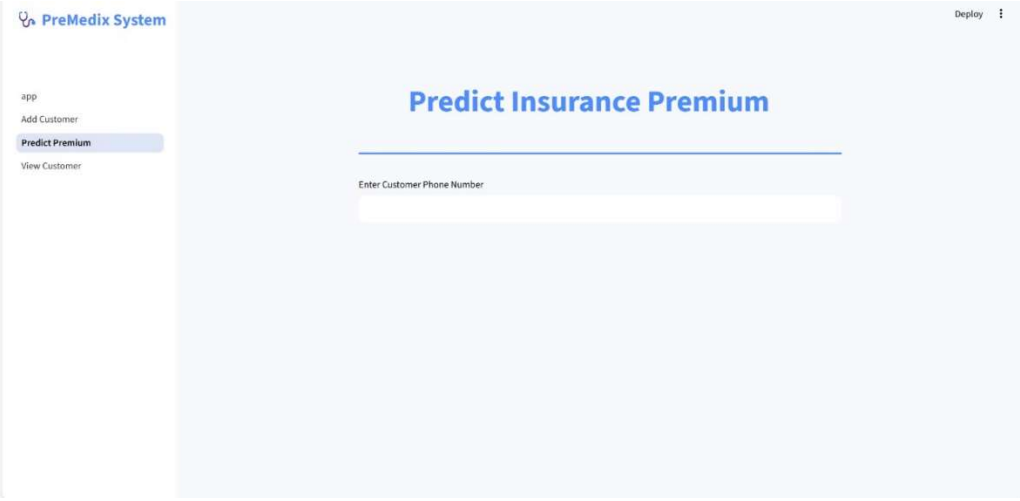
Purpose: This module ensures accurate and structured data collection, forming the foundation for premium prediction.

2. Premium Prediction Module (`_Predict_Premium.py`)

This module integrates the machine learning model to compute insurance premiums for existing customers.

Implementation Highlights

- Loads the saved model (`insurancemodelf.pkl`) using Python's `pickle` library.
- Retrieves customer details by matching phone numbers from the CSV file.
- Automatically extracts required features such as age, BMI, number of children, and smoking status.
- Converts categorical data (e.g., smoker yes/no → 1/0) for model compatibility.
- Passes data into the ML model to generate predictions.
- Assigns a **risk level** based on the predicted premium value:
    - Low
    - Medium
    - High
- Allows saving the predicted premium and risk level back into the customer record.



Purpose: This module transforms raw health data into meaningful insurance insights by leveraging machine learning intelligence.

3. View Customer Module (`_view_Customer.py`)

This module displays all saved customer information in a tabular format for easy analysis.

Primary Implementation Steps

- Reads the `customers.csv` file using pandas.

- Shows customer data in an interactive table using `st.dataframe()`.
- Handles exceptions when the dataset is empty or missing.

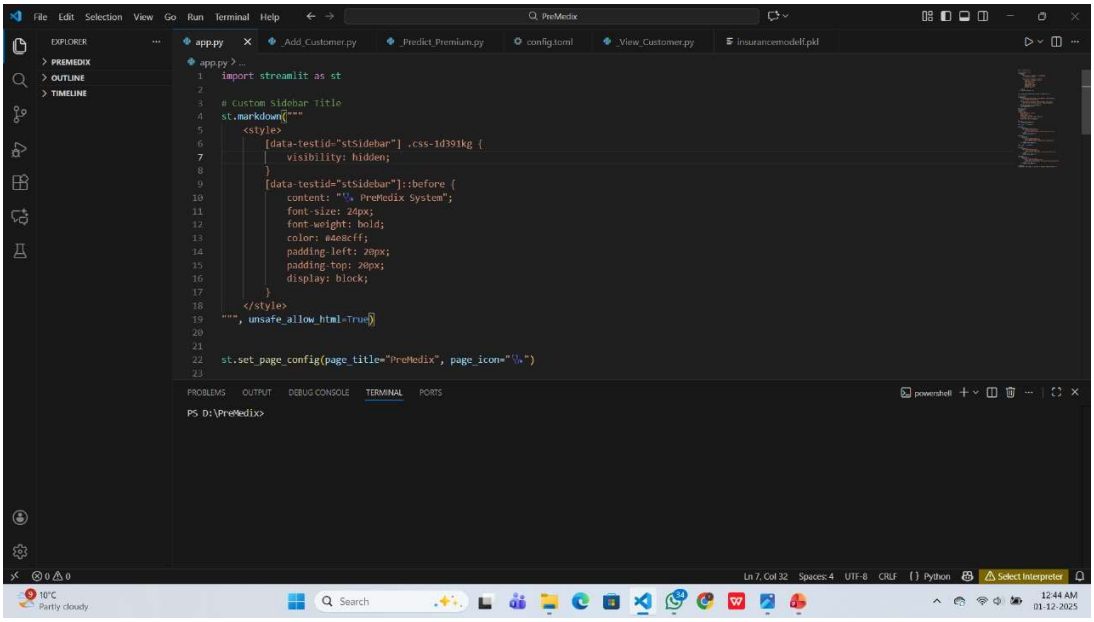
Purpose: This module serves as a centralized dashboard for effortless management and review of customer records.

4. Main Application Interface (`app.py`)

The main application file handles the design, navigation, and user interface styling of the system.

Key Implementation Features

- Sets the global page configuration, title, and icon.
- Adds custom HTML/CSS styling for a modern and professional appearance.
- Displays feature cards such as:
  - Add Customer
  - Predict Premium
  - View Customers
  - Smart Risk Analysis
- Integrates all modules via Streamlit's multi-page structure and sidebar navigation.



Purpose: This file acts as the entry point and enhances the user experience through an attractive and intuitive visual layout.

5. User Interface Theme (`config.toml`)

This configuration file controls the color scheme and typography.

Implementation Details

- Defines primary, secondary, and background colors.
- Sets a modern sans-serif font for clean readability.
- Ensures UI uniformity across all modules.

Purpose: Provides a visually consistent design that improves the overall usability of the application.

6. Machine Learning Model (`insurancemodelf.pkl`)

Although not coded inside the main application, the ML model plays a central role.

Development Overview

- The model is trained on a dataset containing insurance-related attributes.
- Algorithms such as Linear Regression or Random Forest are used.
- The trained model is serialized using `pickle` and loaded into Streamlit for live predictions.

Purpose: Enables fast, accurate, and automated premium estimation.

6.Conclusion and Recommendations

The **PreMedix Insurance System** successfully demonstrates how modern technologies such as Python, Streamlit, and Machine Learning can be combined to create a smart, efficient, and user-friendly medical insurance analysis platform. The system provides an automated solution for collecting customer health details, predicting insurance premiums, and analyzing risk levels.

Through its modular design, the application ensures smooth navigation, accurate data handling, and instant prediction results. The integration of the machine learning model enhances the decision-making process by reducing manual errors and providing objective, data-driven insights. Additionally, the use of a lightweight CSV storage approach makes the system easy to deploy, portable, and practical for small to medium-sized organizations.

Overall, the project meets its objectives by offering:

- A reliable digital method for customer information management
- Accurate insurance premium prediction using ML
- A structured and interactive user interface
- Fast, real-time risk analysis

The successful implementation of the PreMedix system highlights the potential of AI-driven tools in the insurance industry and sets a foundation for future improvements and expansion.

Recommendations

While the system performs effectively in its current form, several enhancements can further improve its performance, scalability, and user experience. The following recommendations are suggested:

1. Integrate a Real Database

Replace CSV files with a database system such as:

- MySQL
- PostgreSQL
- MongoDB

  This will support large-scale data storage, faster retrieval, and multi-user access.

2. Enhance Model Accuracy

Future work may include:

- Training the model on a larger, more diverse dataset
- Trying advanced algorithms (e.g., Gradient Boosting, XGBoost)
- Implementing hyperparameter tuning

3. Add More Health Parameters

Include additional features such as:

- Heart rate
- Cholesterol levels
- Medical history
- Daily activity levels

  More features will improve prediction accuracy and make risk analysis more comprehensive.

4. Implement Authentication & Role-Based Access

Add login systems for:

- Admin
- Agents
- Analysts

  This will ensure data security and controlled access.

5. Add Data Visualization Dashboards

Use charts to display:

- Premium trends
- Risk group distribution
    - Customer demographics

      This will help insurance companies analyze data more effectively.

6. Enable Editing & Deleting Customer Records

Currently, records can only be added and viewed.
Adding editing and deleting features will improve flexibility.

7. Deploy the Application Online

Deploying on cloud platforms like:

- Streamlit Cloud
- AWS
- Heroku

  will make the system accessible to insurance employees from anywhere.

7.References

- **Streamlit Documentation** – Official guide for building web apps in Python.
https://docs.streamlit.io/

- **Pandas Library Documentation** – Used for data handling and CSV operations.
https://pandas.pydata.org/

- **Python Official Documentation** – Core language reference for implementation.
https://www.python.org/doc/

- **Scikit-learn Documentation** – Used for training and loading the machine learning model.
https://scikit-learn.org/

- **Insurance Dataset Reference** – Dataset commonly used for insurance premium prediction.
https://www.kaggle.com/mirichoi0218/insurance

Github:-
https://github.com/rosyrandhawa/ROSY_12200724_B49A_2_/