**Chronic Disease Risk Prediction and Personalized Insurance Recommendation System**

**Introduction**

This report outlines the design, development, and implementation of a personalized insurance recommendation system. The project integrates structured and unstructured data, machine learning predictions, and a user-friendly interface to deliver tailored insurance recommendations. The system aims to improve user experience, enhance operational efficiency, and provide actionable insights for insurance providers.

**Objectives**

* Improve User Experience: Capture user health metrics and provide personalized recommendations.
* Integrate Machine Learning: Use predictive models to assess chronic disease risks.
* Enhance Data Processing: Combine structured and unstructured data for deeper insights.

**Implementation Details**

Technologies Used

* Database: MySQL for structured data management.
* Machine Learning: Scikit-learn for model training and prediction.
* Interface: Python-based CLI for user interaction.

Key Functions

* Data Fetching and Updating: Optimized queries for fetching and updating user metrics.
* Unstructured Data Processing: Sentiment analysis to extract meaningful insights from user inputs.
* Prediction Updates: Only predict and update data for users with new metrics.

**System Workflow**

* User Interaction
  + The system begins by engaging users in a **conversational interface** where they answer health-related questions. These responses are stored as **unstructured data** for further analysis.
* Data Analysis and Storage
  + Responses are analyzed to derive sentiment scores for mental, physical, and emotional well-being.
  + Sentiment scores are stored in the HealthMetrics table, enriching the structured data.
* Machine Learning Predictions
  + A RandomForestClassifier predicts chronic disease risk based on health metrics and sentiment scores.
  + Predictions are updated in the ChronicDiseaseRisk table, providing actionable insights.
* Recommendation Generation
  + Based on the predicted risk level, the system identifies and recommends suitable insurance products.

**Database Schema**

Key Tables

* Customer: Stores user demographic information.
* HealthMetrics: Contains both structured health data and derived sentiment scores.
* ChronicDiseaseRisk: Stores ML predictions for chronic disease risks.
* UserNotes: Captures unstructured data from user interactions.
* Product: Contains details of available insurance plans.

**Machine Learning Integration**

Model Selection

* The system uses a RandomForestClassifier to predict chronic disease risks. This model was chosen for its robustness and ability to handle mixed data types.

Features

* Key features include:
  + Traditional health metrics: BMI, Heart Rate, Blood Pressure, etc.
  + Sentiment scores from unstructured data: Mental, Physical, Happiness.

Achievements from Part 3

* System Design:
  + Database schema includes normalized tables for customer information, health metrics, chronic disease history, and products.
  + The Product table has a risklevel column mapped to predicted risk levels for product recommendations.
* Data Generation:
  + Randomized data simulates real-world scenarios, categorizing clients into diagnosed, undiagnosed, and undetermined groups.
  + Health metrics and product details are generated for testing and demonstration purposes.
* Machine Learning Logic:
  + A Random Forest model predicts chronic disease risk, scaling results from 0-100 to 0-5 for easier mapping.
* Recommendation Logic:
  + The system fetches a random undetermined client using the get\_random\_undetermined\_client function.
  + Based on the client’s risk level, a product is recommended via the recommend\_products function.
  + Personalized messages include the client’s name and the recommended product.

**Business Use Case**

A diagram of a software process

Description automatically generated

The personalized insurance recommendation system significantly enhances customer engagement, operational efficiency, and decision-making for insurance providers. By integrating both structured data (e.g., health metrics, demographic information) and unstructured data (e.g., user inputs on mental, physical, and emotional states), the system enables personalized interactions that foster trust and transparency.

From a customer’s perspective, the system offers an intuitive and conversational interface, making it easy to provide their health information and receive tailored insurance recommendations. The sentiment analysis of unstructured data, combined with machine learning predictions, ensures that recommendations are accurate and relevant.

For insurance providers, the system streamlines the workflow by automating data collection, analysis, and product recommendation. The integration of machine learning not only improves risk assessment but also optimizes the alignment between risk levels and product offerings, enabling providers to make informed decisions and deliver better customer service. Additionally, the system’s ability to dynamically update predictions and metrics ensures that the recommendations remain up-to-date and actionable, creating a seamless user experience.

**Future Opportunities**

* Scalability
  + Expand the system to handle larger datasets and more complex workflows.
* Frontend Development
  + Develop a web-based or mobile application to enhance user interaction.
* Advanced Analytics
  + Incorporate NLP techniques for deeper analysis of unstructured data.
* Proactive Features
  + Notify users of significant changes in health risks or new insurance plans.

**Conclusion**

This personalized insurance recommendation system successfully integrates structured and unstructured data with machine learning to deliver value to both users and insurance providers. The project lays a foundation for future enhancements in user engagement and predictive analytics.

**Example Usage (FINAL):**

A screenshot of a computer screen

Description automatically generated

**Example Usage (TESTING):**

**A screen shot of text

Description automatically generated**

Before Sentiment Analysis:

* The system initially provides a recommendation for the user (Daniel Black) based on existing health metrics and chronic disease risk predictions.
* In this case, the original recommendation is generated based on the user's pre-existing health data and the model's earlier predictions.

New Data is Stored and Updated:

* The system engages the user in a conversational interface, asking questions about their mental, physical, and emotional state:
  + Mental: “good”.
  + Physical: “super food”.
  + Happiness: “perfect”.
* These responses are stored as unstructured data in the UserNotes table and analyzed to generate sentiment scores.
* The updated sentiment scores are then stored in the HealthMetrics table.

New Predictions are Made:

* Using the enriched health metrics (including the sentiment scores), the machine learning model recalculates the chronic disease risk for the user.
* The new risk level and confidence score are updated in the ChronicDiseaseRisk table, reflecting the latest data.

New Recommendation:

* Based on the updated predictions, the system provides a new insurance recommendation, which better aligns with the user's updated health profile and risk assessment.