Health Assistant Using Machine Learning

HALDIA INSTITUTE OF TECHNOLOGY

Dept : INFORMATION TECHNOLOGY

Submitted To :- Mr. Debasish Sahoo

Submitted By :--

Name	Roll No	Email
Subhadip Dey	10300223170	deysubhadip66@gmail.com
Srijani Dinda	10300223169	srijanidinda64@gmail.com
Sathi Pramanik	10300223137	18sathipramanik@gmail.com

Problem Statement

- Doctors and healthcare staff are required to make **quick and critical decisions** based on multiple patient parameters such as age, vitals, medical history, and lifestyle factors.
- **Manual analysis of patient records** is time-consuming and may not always provide accurate or timely insights, especially when dealing with large volumes of data.
- Predicting **disease conditions, recovery duration, treatment cost, and readmission risk** becomes challenging due to the complexity and size of healthcare datasets.
- The lack of automated data analysis can lead to **delayed diagnosis, inefficient treatment planning, and increased workload** for healthcare professionals.

Project Objectives

1

Disease Prediction

Predict possible diseases using patient health parameters and vital signs through machine learning algorithms

2

Risk Assessment

Estimate individual health risk levels and potential recovery duration based on predicted conditions

3

Personalized Guidance

Provide personalized, educational health suggestions through an intelligent assistant system

Machine Learning Methodology

Training Approach Approach

Supervised machine learning models trained on validated hospital patient datasets

Feature Analysis

Feature-based analysis using patient vitals, medical history, and demographic information

Disease Clustering

Advanced disease clustering with probability-based ranking algorithms

Prediction Output

Confidence-oriented prediction results with statistical reliability metrics

Dataset Overview

- **Total Records:** 10,000 patient records, representing a large and diverse sample suitable for reliable statistical analysis and machine learning model training.
- **Total Attributes:** 28 features covering patient demographics, physical measurements, health indicators, hospital operations, financial information, and medical outcomes.
- **Data Types:** A combination of **numerical data** (such as age, vital signs, hospital stay duration, and treatment cost) and **categorical data** (such as gender, disease type, lifestyle habits, insurance status, and treatment outcomes).
- **Data Nature:** The dataset simulates a **real-world hospital environment**, reflecting practical healthcare scenarios and patient variability.
- **Use Cases:** Designed for **healthcare analytics**, including disease prediction, treatment cost estimation, patient outcome analysis, and hospital readmission risk assessment using machine learning techniques.

Dataset Attributes

Patient Demographic Details
Patient_ID , Patient_Name , Age , Gender , City

Physical Details
Height_cm , Weight_kg

Health Metrics
Pulse_Rate , Oxygen_Level ,BP ,
Sugar ,Chol

Lifestyle & Risk Factors
Smoke, Alcohol , Stress_Level ,
Chronic_Disease

Hospital & Treatment Details
Doctor_Specialization, Appointment_Type,
Admit_Type, ICU_Required, Days, Visits

Medical Outcomes
Disease , Outcome, Readmitted, Status

Financial Information
Cost , Insurance

Feature Selection

To improve model accuracy and reduce complexity, the most relevant health-related features were selected based on correlation analysis and domain knowledge.

➢ Selected Features

- ❑ **Age** – Influences disease risk and treatment cost
- ❑ **Height (cm)** – Used for physical assessment and BMI calculation
- ❑ **Weight (kg)** – Indicates body condition and health risk
- ❑ **Blood Pressure (BP)** – Key indicator of cardiovascular health
- ❑ **Sugar Level** – Important for diabetes and metabolic conditions
- ❑ **Cholesterol Level** – Associated with heart disease risk
- ❑ **Days in Hospital** – Reflects severity of illness and treatment duration
- ❑ **Number of Visits** – Indicates frequency of medical attention required
- ❑ **Treatment Cost** – Primary target variable for prediction

Machine Learning Models Used

Classification Models

Random Forest Classifier

- Ensemble-based model using multiple decision trees
- Handles complex relationships and reduces overfitting
- Provides high accuracy and feature importance

K-Nearest Neighbors (KNN)

- Distance-based algorithm
- Classifies patients based on similarity to nearby data points
- Useful for pattern recognition in patient health profiles

Logistic Regression

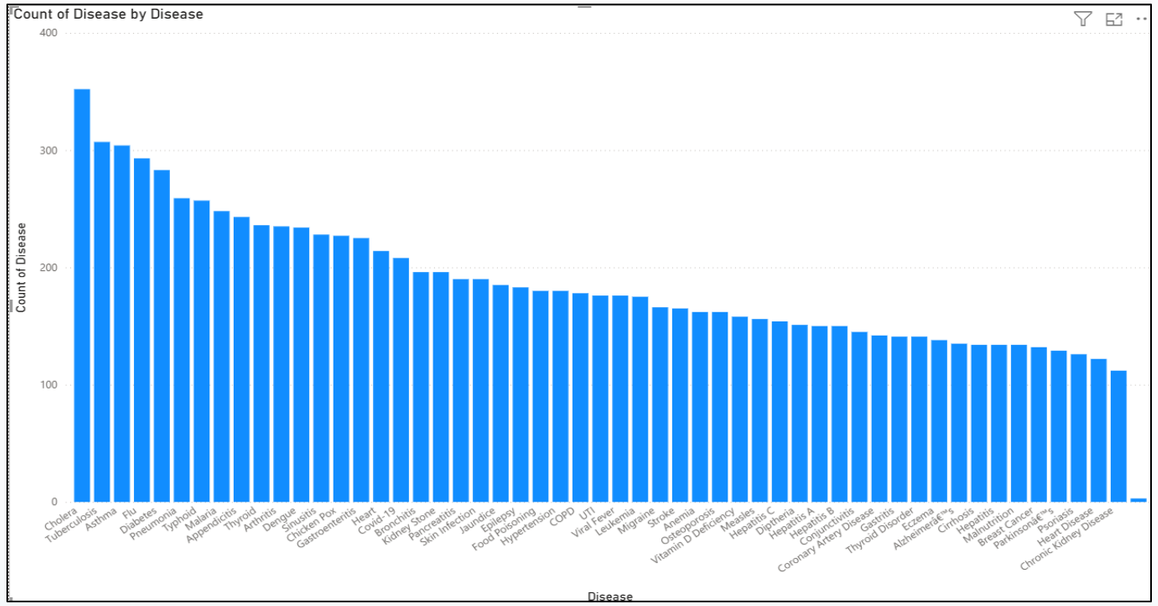
- Simple and interpretable linear classification model
- Effective for binary outcomes such as *Recovered / Not Recovered*
- Works well with well-selected features

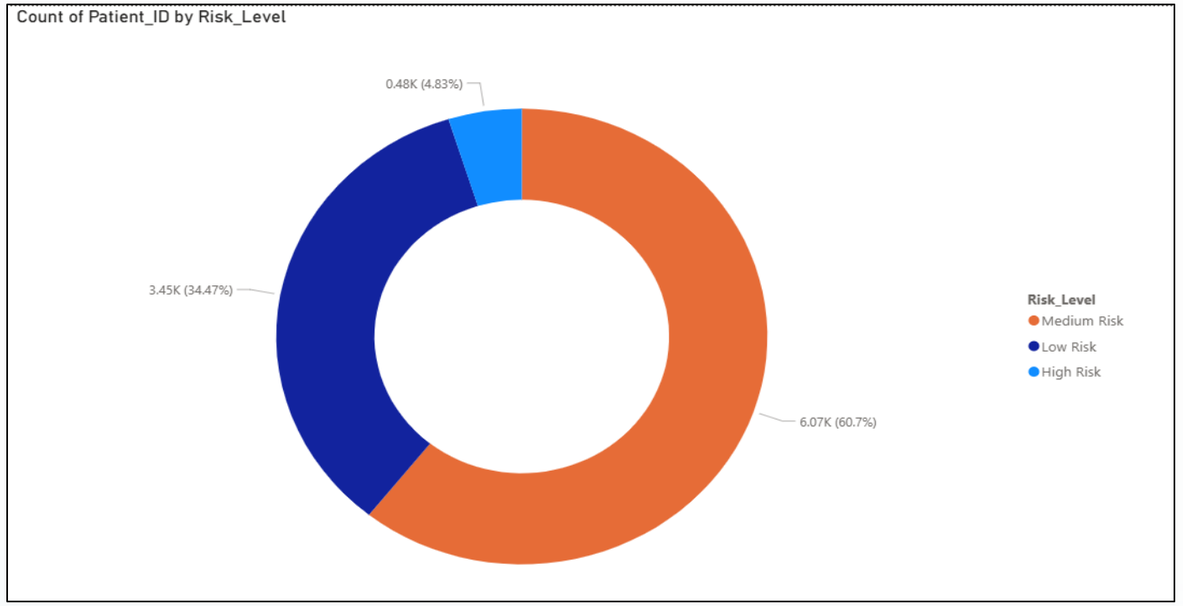
Regression Model

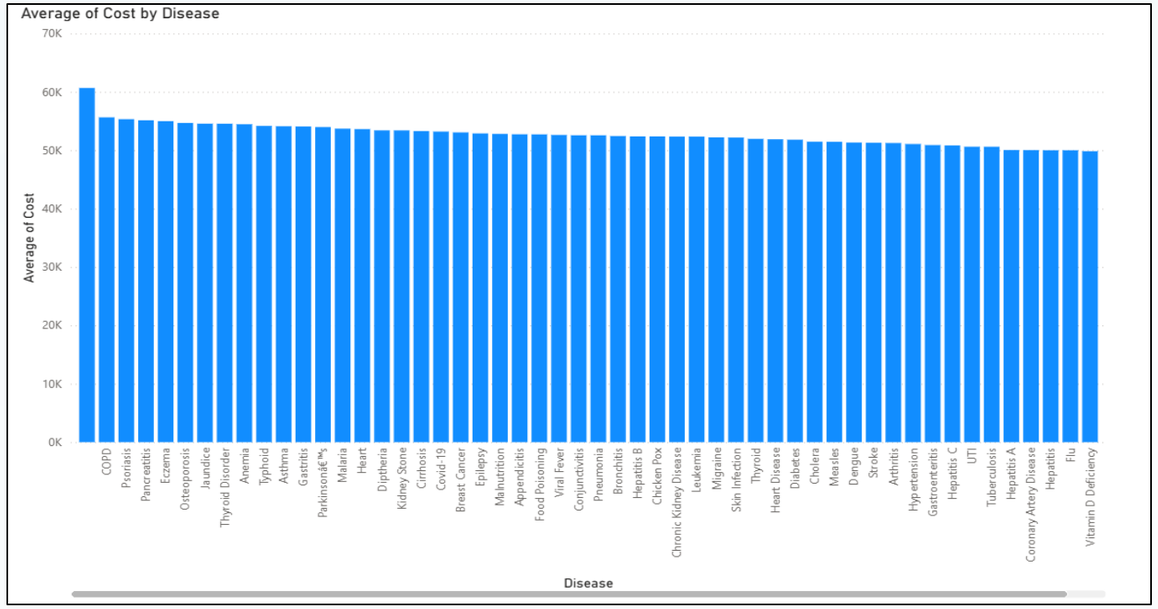
Random Forest Regressor

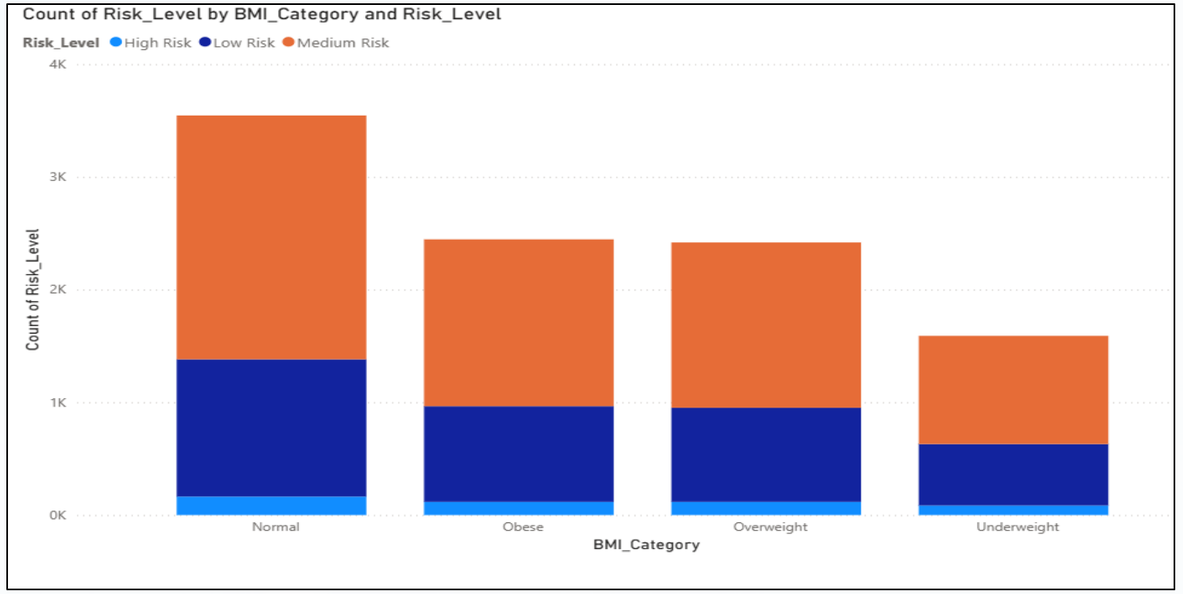
- Predicts the number of days required for patient recovery
- Captures non-linear relationships between health parameters
- Robust to noise and performs well on real-world healthcare data

Data Analysis Report









Core Functionalities

Disease Prediction

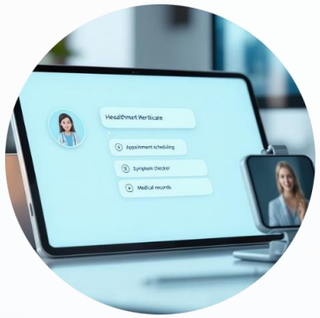
Generates disease predictions with confidence scores and probability rankings

BMI & Risk Classification Classification

Calculates body mass index and classifies individual health risk levels

Top Disease Display

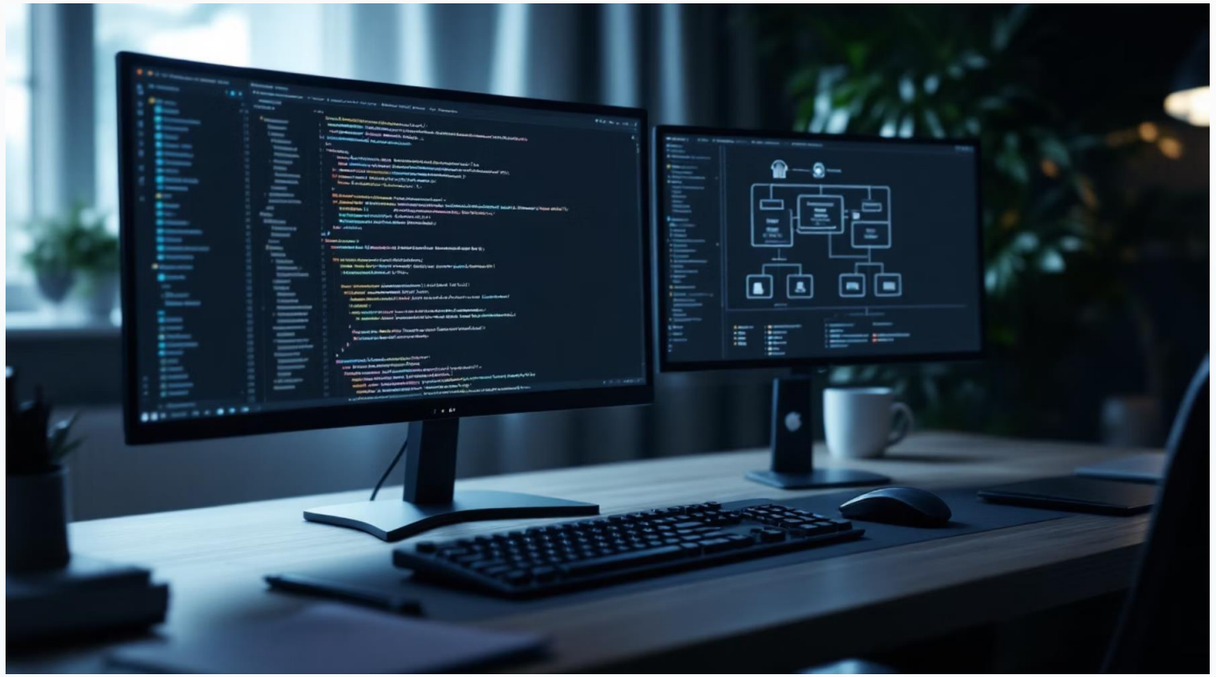
Displays top three probable diseases ranked by statistical likelihood



Health Query Assistance Assistance

Provides context-aware responses to user health questions and concerns

Technology Stack



Frontend

React.js framework for responsive, interactive user interface design

Backend

Python with Flask framework providing RESTful API services

Machine Learning

Scikit-learn library for model development and prediction algorithms

Data Processing

Pandas and NumPy libraries for efficient data manipulation and analysis

Health Assistant

Local large language model (LLM) deployed using Ollama platform

Results and Advantages

Reliable Prediction

Delivers accurate disease prediction based on comprehensive patient data analysis

Health Awareness

Improves patient health awareness and understanding of potential risk factors

User-Friendly Interface

Provides intuitive and responsive web interface accessible across devices

Privacy-Focused

Ensures intelligent assistance while maintaining strict patient data privacy standards

Conclusion

The ML-Based Health Prediction and Assistance System demonstrates the practical application of machine learning in healthcare, providing accessible, data-driven health insights while maintaining user privacy and system reliability.

> This project represents a step forward in making preliminary health assessment more accessible and empowering individuals with intelligent, personalized health information.

THANK YOU