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

Introduction

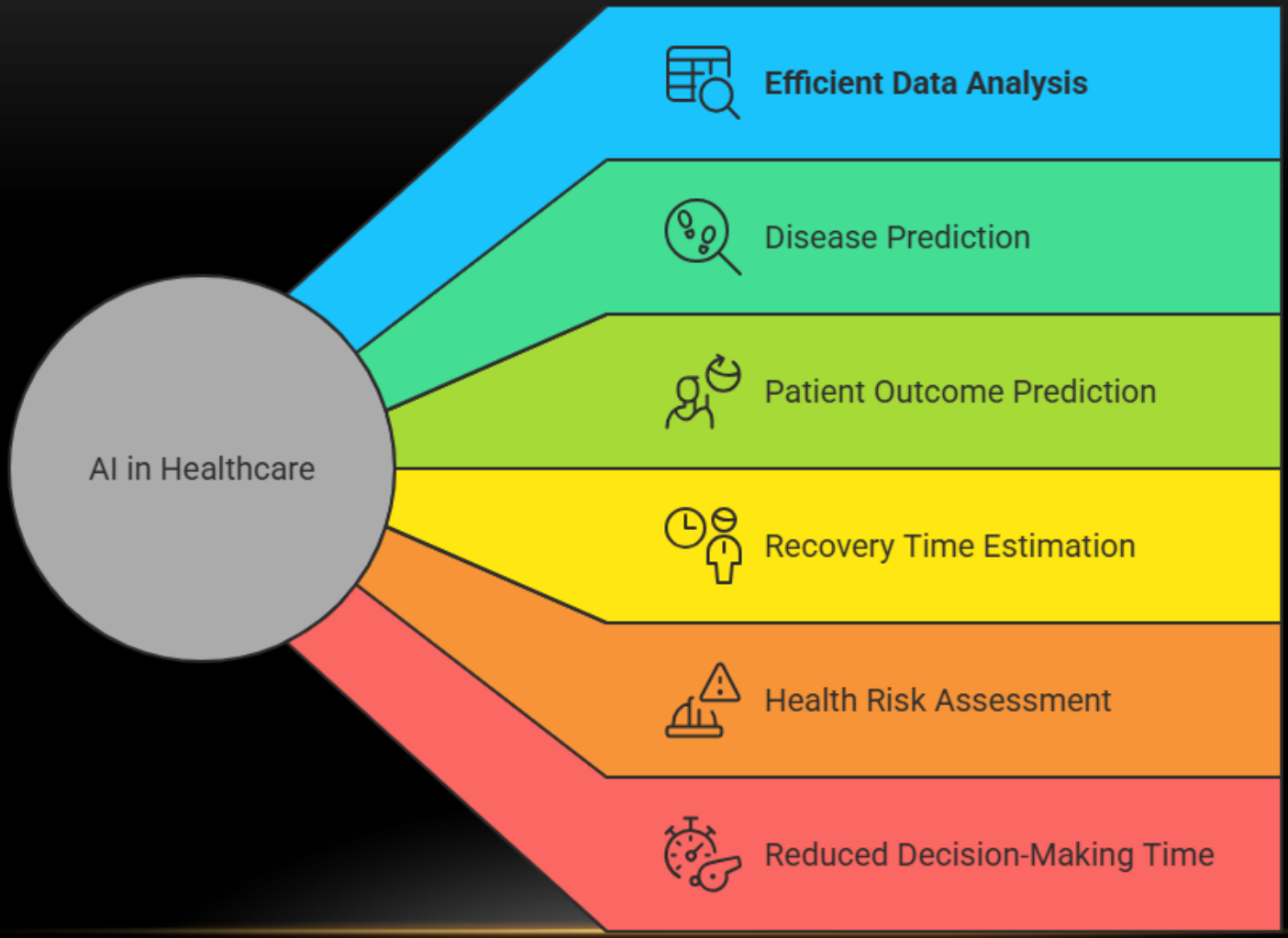
In today's rapidly evolving healthcare environment, hospitals generate a massive volume of patient data on a daily basis. This data includes medical history, laboratory test results, vital signs, lifestyle factors, and treatment records. Managing and analyzing such large and complex datasets manually is not only time-consuming but also highly prone to human error.

Machine Learning (ML) provides an effective solution by enabling automated analysis of healthcare data to uncover hidden patterns, trends, and relationships. By leveraging ML techniques, healthcare systems can assist doctors in early disease detection, accurate diagnosis, personalized treatment planning, and risk assessment. This data-driven approach enhances clinical decision-making, reduces workload on medical professionals, and improves overall patient care and operational efficiency.

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.

Objectives


AI in Healthcare
Efficient Data Analysis
Disease Prediction
Patient Outcome Prediction
Recovery Time Estimation
Health Risk Assessment
Reduced Decision-Making Time

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

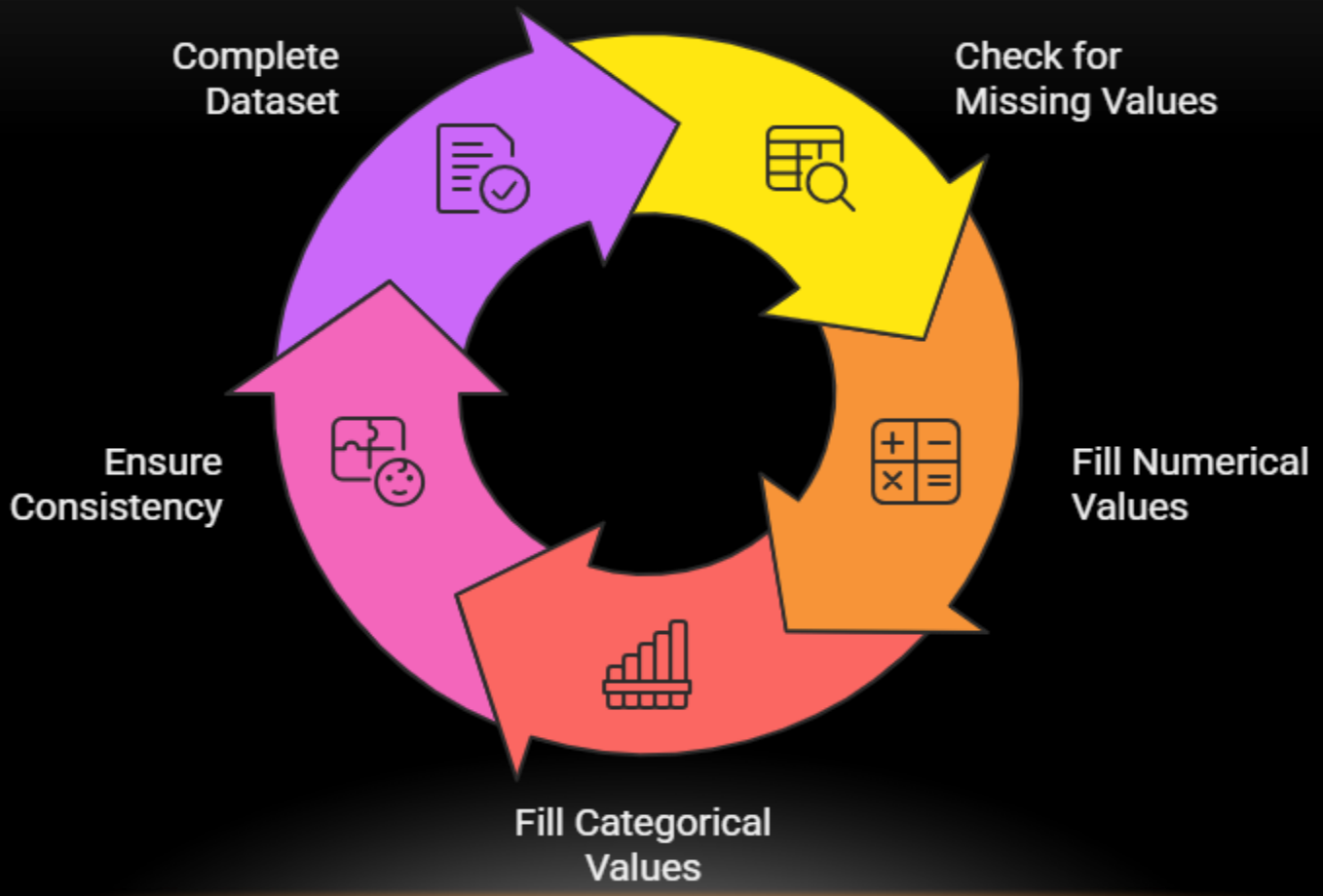
Lifestyle & Risk Factors
Smoke, Alcohol , Stress_Level , Chronic_Disease

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

Medical Outcomes
Disease , Outcome, Readmitted, Status

Financial Information
Cost , Insurance

Data Preprocessing



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

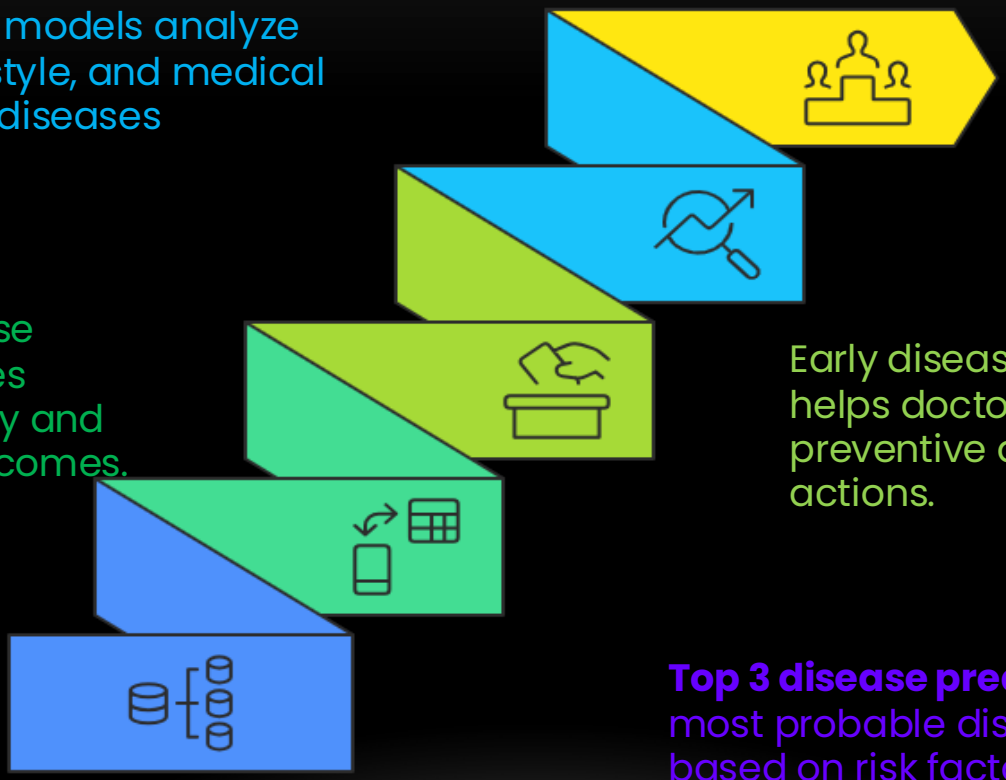
Random Forest Regressor

- 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

Disease Prediction

Machine learning models analyze patient vitals, lifestyle, and medical history to predict diseases accurately.

Data-driven disease prediction improves diagnosis accuracy and overall patient outcomes.



Automated prediction reduces manual effort and supports faster clinical decision-making.

Early disease prediction helps doctors take timely preventive and treatment actions.

Top 3 disease prediction identifies the most probable diseases for a patient based on risk factors.

Health Risk Assessment

- Health risk calculated using:

  - Blood Pressure
  - Sugar level
  - Cholesterol
  - Age

- Risk categories:

  - Low Risk
  - Medium Risk
  - High Risk

- Helps in early warning of critical patients

System Output

01 Predict Disease

02 Evaluate Outcome

03 Monitor Health Status

04 Assess Readmission

05 Estimate Recovery Days

06 Determine Health Risk

Tools & Applications

Programming Language

Python and Flask for backend, ReactJS for frontend.

Libraries

Pandas, NumPy, and Scikit-learn for data analysis.

Platform

Google Colab and VSCode for development.

Model Saving

Joblib for saving and loading machine learning models.

Applications

Hospital decision support, patient monitoring, and healthcare analytics.

Conclusion & Future Scope

- ❑ Successfully developed ML-based Health Assistant
- ❑ Demonstrates practical ML application in healthcare
- ❑ Provides valuable health insights
- ❑ Can support medical professionals effectively

Future Scope:

- Use real-time hospital data
- Improve accuracy with advanced ML
- Web or mobile application integration

THANK YOU