Project Report

Project Title: Leveraging Machine Learning for the Early Detection of Hypertension Through Routine Health Parameters.

Introduction

Hypertension, commonly known as high blood pressure, is one of the most prevalent non-communicable diseases globally and a leading contributor to cardiovascular morbidity and mortality rate. According to the World Health Organization (WHO), approximately 1.28 billion adults aged 30-79 years globally have hypertension, and a significant proportion of them are unaware of their condition. Hypertension may not present obvious symptoms until severe damage to vital organs such as heart, brain or kidneys as occurred, it is often referred to as the "silent killer". Early detection and management of individuals at risk can help in timely intervention and prevention. This project aims to develop a machine learning-based web application that predicts the risk of hypertension based on user-provided health data.

Problem statement

Hypertension is a major risk factor for cardiovascular diseases such as heart attacks, strokes and kidney failure. Despite its serious implications, it often goes undiagnosed in its early stages because it may not present noticeable symptoms. In many low-resource setting, access to specialized diagnostic tools and regular checkup is limited, leading to late detection and poor health outcomes.

Routine health parameters such as age, sex, blood pressure readings, weight, height, body mass index (BMI) etc are readily made available during basic health assessments. However, these parameters are underutilized in proactive health screening. There is a growing need for cost effective, accessible and accurate tools that can leverage these routine metrics to identify individuals at risk of hypertension early.

This project aims to bride this gap by developing a machine learning model that can analyze routine clinical data and accurately predict an individual's risk of hypertension. The goal is to empower healthcare providers with a simple yet powerful decision-support tool to enhance early diagnosis and timely intervention.

Aim

To develop a user-friendly and easily accessible machine learning-based web application that predicts an individual's risk of hypertension based on their routine health parameters.

Objectives

- To collect and preprocess a health-related data including age, sex, blood pressure, cholesterol, ecg etc.
- To build a predictive model that can classify individuals as hypertensive or non-hypertensive.
- To deploy this model using a user-friendly web interface.
- To demonstrate the integration of machine learning with real-time web-based predictions.

Dataset

The dataset used was sourced from kaggle, a public health-related dataset repository. It contains several clinical and lifestyle-related features such as:

- Age
- Sex
- Chest pain type (cp)
- Resting blood pressure (trestbps)
- Cholesterol level (chol)
- Fasting blood sugar (fbs)
- Resting electrocardiographic results (restecg)
- Maximum heart rate achieved (thalach)
- Exercise-induced angina (exang)
- ST depression (oldpeak)
- Slope of the ST segment (slope)
- Number of major vessels colored by fluoroscopy (ca)
- Thalassemia (thal)
- Outcome (used as dummy input for consistency)

Model Development

- **Algorithm Used**: Logistic Regression, Random Forest and XGBoost Classifier.
- **Preprocessing**:

i. Feature scaling using StandardScaler.
ii. Handling missing values by filling them with 0.

- **Training**: Dataset was split into training and test sets (e.g 80/20).
- Model evaluation metrics included accuracy, precision, recall, and confusion matrix.

Web Application

The model was deployed using Flask, a lightweight Python web framework. Key technologies includes:

- **Backend**: Flask, joblib (for model loading), pandas, sklearn.
- **Frontend**: HTML, CSS (with custom styling), JavaScript (for API communication).
- **Prediction Flow**:

i. User inputs health-related data into the form.
ii. Data is sent via POST request to the Flask backend.
iii. The model processes the input and returns a prediction.
iv. The user interface (UI) displays the risk level (Low or High) to the user.

Features

- Clean, responsive user interface.
- Supports full set of model-required features.
- Auto-handles missing fields by defaulting them to 0.
- Visual feedback with animations and styling.
- Error handling for server or prediction issues.

Results

- The best-performing model achieved an accuracy of approximately 85%.
- The web app was tested with both high-risk and low-risk examples and returned reliable predictions.
- Input validation ensures only realistic values are accepted.

Limitations

- The current model assumes a static set of features and default mappings.
- The input data used may not reflect all ethnic or demographic variances.
- Outcome feature is currently hard-coded to ensure model compatibility.

Future Improvements

- Collect real-world data for validation.
- Add support for BMI and lifestyle insights like stress and diet.
- Integrate user authentication for saving prediction history.