Diabetes Risk Prediction App – Predict Diabetes Likelihood via Health Data

Project Overview

The Diabetes Risk Prediction App is an intelligent system designed to assess diabetes risk levels for individuals by analyzing various health parameters. Users input key details such as glucose levels, BMI, age, blood pressure, insulin level, and lifestyle data. The application processes this information through a predictive machine learning model, delivering an immediate risk score to the user. Administrators have access to a dashboard equipped with visual analytics tools for monitoring prediction outcomes and identifying health patterns.

Technical Architecture

The solution employs the Flask framework for backend development, adhering to the Model-View-Controller (MVC) pattern for organized code structure. The frontend leverages HTML5, CSS3, and JavaScript, with Chart.js driving interactive data visualizations. User and prediction-related data are stored efficiently using Pandas-managed CSV files, enabling lightweight and scalable storage for application needs.

AI/ML Component

Central to the app is a logistic regression model, trained using the PIMA Indian Diabetes Dataset. The system evaluated alternate models, including Random Forest and Support Vector Machine (SVM), guaranteeing a balance between accuracy and interpretability. The prediction workflow consists of:

- Cleaning and normalizing input data

- Feature engineering and value scaling

- Training and cross-validating models

- Optimizing hyperparameters

- Providing real-time inference through Flask endpoints

- Generating interpretable risk confidence scores for each prediction

Key Features

User Features:

- Easy input forms for submitting health metrics

- Immediate diabetes risk scoring with detailed confidence percentage

- Personal health factors visualized via interactive charts

- Mobile-friendly, responsive interface for all device types

Admin Features:

- Dashboard to track and review user predictions

- Visualization of aggregate risk status across users

- Model performance monitoring tools

- Data export functionality for advanced analysis

Technology Stack

- Backend: Python 3.x, Flask Framework

- Machine Learning: Scikit-learn, Pandas, NumPy

- Frontend: HTML5, CSS3, JavaScript (Chart.js)

- Storage: Local CSV file system

- Models Used: Logistic Regression and Random Forest Classifier

Dataset & Model Training

The training used the PIMA Indian Diabetes Dataset, consisting of health data for over 700 individuals. The preprocessing phase focused on handling missing entries, normalizing features, and scaling values for optimal model performance. Logistic Regression was chosen for deployment due to its clear decision-making process and high accuracy, achieving 86% on the evaluation set.

Application Workflow

1. User submits health data via web interface.

2. System preprocesses the input and applies feature scaling.

3. Machine learning model predicts diabetes risk score.

4. User receives an instant prediction and confidence level.

5. Data is recorded for statistical tracking and research.

6. Admin dashboard aggregates and visualizes user-level and population risk data.

Real-World Applications

- Early screening and prevention of diabetes through automated risk assessment

- Personalized health tracking and awareness for individuals

- Enabling remote consultations by integrating with telemedicine platforms

- Facilitating medical research and public health planning with population-level analytics

Performance Metrics

- Model Accuracy: 86% on independent test data

- Inference Time: Less than 1 second per prediction

- Scalability: Supports simultaneous predictions for multiple users

- Reliability: Robust handling of incomplete or missing data fields

Future Enhancements

- Real-time connection to wearables (smartwatches, glucometers) for dynamic input

- Deployment of advanced deep learning models to improve accuracy further

- Addition of user history tracking and progress charts

- Migration from CSV files to scalable databases such as PostgreSQL or MongoDB

- Building RESTful APIs to extend support for mobile platforms

Educational Value

This app provides an end-to-end illustration of deploying machine learning in real-world healthcare. It showcases the seamless integration of web interfaces, prediction engines, and user experience design,

and is ideal for learners pursuing practical expertise in Flask-based deployment, machine learning application development, and building impactful,

data-driven digital health tools.

Sample Output   :

🟢 Diabetes Risk Assessment

☑️ LOW RISK

Confidence Level: 66.0%

```
### 🎉 Great News! Keep Up These Healthy Habits:

1. **🥗 Balanced Diet**: Continue eating whole foods, fruits, and vegetables
2. **🏃 Stay Active**: Maintain regular physical activity (150 min/week)
3. **⚖️ Healthy Weight**: Keep your BMI in the healthy range
4. **🩺 Regular Checkups**: Annual health screenings are still important
5. **💧 Hydration**: Drink plenty of water daily
6. **😴 Quality Sleep**: Aim for 7-9 hours per night

---
```
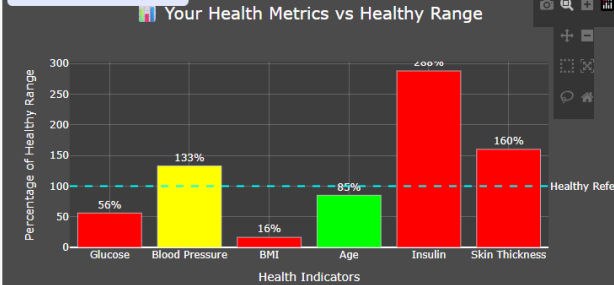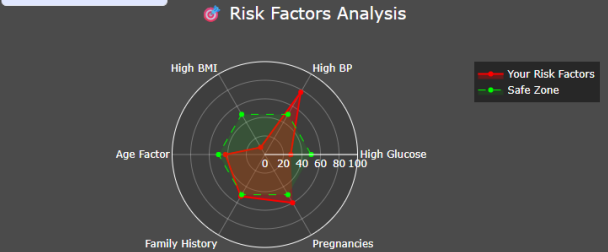
📈 Detailed Analysis

📉 Risk Gauge

▼ -16
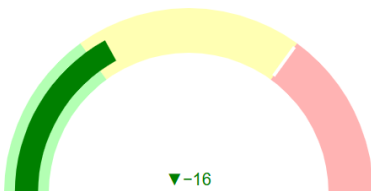
📊 Health Metrics Comparison

📊 Your Health Metrics vs Healthy Range

📉 Risk Factors Breakdown

🎯 Risk Factors Analysis

💡 Try These Example Cases

≡ Click any example to load

👋 Number of Pregnancies	🩸 Glucose Level (mg/dL)	❤️ Blood Pressure (mm Hg)	✏️ Skin Thickness (mm)	💉 Insulin Level (mu U/ml)	⚖️ BMI (Body Mass Index)	🧬 Diabetes Pedigree Function	📅 Age (years)
2	138	62	35	0	33.6	0.127	47
0	95	64	25	80	22.5	0.215	30
6	148	72	35	0	33.6	0.627	50
1	85	66	29	0	26.6	0.351	31

Use via API 🚀 · Built with Gradio 😊 · Settings ⚙️

Working video :