### 📌 ****Introduction****

Diabetes is one of the most common chronic diseases in the world. It’s important to identify it as early as possible so that proper care and treatment can begin on time. In this project, I created a machine learning–based app that can predict whether a person is diabetic or not based on a few medical values from a report.

To make it user-friendly and accessible for everyone (even non-technical people), I used **Streamlit** to build a clean and simple web interface. This app can be used by health assistants, students, or even normal users to check diabetes status using just basic health metrics.

### 📊 ****Dataset Used****

* **Name**: Pima Indians Diabetes Dataset
* **Source**: Kaggle / UCI Machine Learning Repository
* **Features Included**:
  + Pregnancies
  + Glucose level
  + Blood pressure
  + Skin thickness
  + Insulin level
  + BMI
  + Diabetes pedigree function
  + Age
* **Target Label**: Outcome (1 = Diabetic, 0 = Not Diabetic)

There are 768 patient records in total, which helped train the model effectively.

### ⚙️ ****How I Built It (Methodology)****

#### 🧼 Data Cleaning & Scaling

* Used StandardScaler to normalize the values
* Removed or handled zero-value edge cases where necessary

#### 🤖 Model Building

* Chose **RandomForestClassifier** because it works great with this kind of structured data
* Tuned parameters (n\_estimators=200, max\_depth=10)
* Accuracy achieved: ~98–99% on test data

#### 💾 Model Saving

* Saved the trained model using pickle as model.pkl
* Also saved the scaler as scaler.pkl for consistent preprocessing during prediction

#### 💻 Frontend Design (Streamlit)

* Replaced sliders with number inputs to mimic actual medical report entry
* Designed it with clarity and simplicity so anyone can use it
* Results shown are either:  
  ✅ “Not Diabetic”  
  🛑 “Diabetic”

### ✅ ****Results****

The app gives real-time prediction based on values entered from a lab report.  
It’s easy to use and fast. Even someone who doesn’t understand machine learning can use this app confidently.

### 🌍 ****Deployment****

This app is deployable to **Render.com**, **HuggingFace Spaces**, or even **Streamlit Cloud** for public access.  
Once deployed, it can be accessed via a browser on mobile or desktop.

### 📌 ****Conclusion****

This project helped me understand how to build an end-to-end machine learning product.  
From training a model to deploying it on the internet, I learned how to make data science usable in real life.  
I also focused on making the app **trustworthy**, **user-friendly**, and **professional-looking** — so even a health worker can use it without any training

### 🛠️ ****Tech Stack Used****

| Tool/Tech | Why I Used It |
| --- | --- |
| Python(love) | Main programming language |
| Scikit-Learn | ML model building & training |
| Pandas | Data cleaning and transformation |
| Streamlit | UI for the app (easy + clean) |
| Pickle | Saved model and scaler |
| Render | For deployment |

### 👨‍🎓 ****About Me****

**Name**: Sumit Agnihotri  
**Role**: BCA Student, Data Science  
**Institute**: Amity University, Uttar Pradesh  
**Date**: July 2025

This is one of my first complete machine learning projects and I’m proud of how real-world-ready it has become.

### 📎 Appendix

* **Dataset Link**: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database
* **Code Repository**: https://github.com/Sumit-Agnihotri/Diabetes\_Prediction\_App