Diabetes Prediction Project - Complete Documentation

# 1. Project Overview

This project is an end-to-end machine learning solution to predict whether a person has diabetes based on health metrics. It uses the Pima Indians Diabetes Dataset and includes model training, Flask-based web deployment, and an optional Streamlit GUI.

# 2. Project Folder Structure

diabetes\_prediction\_project/  
│  
├── model/ # Trained model and scaler (.pkl files)  
│ ├── diabetes\_model.pkl  
│ └── scaler.pkl  
├── templates/ # HTML frontend (for Flask)  
│ └── index.html  
├── app.py # Flask backend  
├── streamlit\_app.py # Optional Streamlit frontend  
├── train\_diabetes\_model.ipynb # Jupyter notebook for training  
├── requirements.txt # Python dependencies  
├── render.yaml # Deployment config for Render  
└── diabetes.csv # Dataset (downloaded manually from Kaggle)

# 3. Model Training (train\_diabetes\_model.ipynb)

The Jupyter Notebook handles:  
- Loading the dataset  
- Handling missing values  
- Splitting and scaling data  
- Training a Random Forest model  
- Saving the model and scaler

# 4. Flask Web App (app.py)

The Flask app loads the trained model and scaler, exposes two routes:  
- `/` for the HTML form  
- `/predict\_form` for handling form predictions

# 5. HTML Frontend (templates/index.html)

The HTML form captures 8 input features and posts them to the Flask backend for prediction.

# 6. Optional Streamlit App (streamlit\_app.py)

This version uses Streamlit for a cleaner and simpler user interface. Run it with:  
streamlit run streamlit\_app.py

# 7. Running the Project

Step 1: Set up virtual environment  
----------------------------------  
python -m venv venv  
venv\Scripts\activate  
pip install -r requirements.txt  
  
Step 2: Train the model  
-----------------------  
jupyter notebook  
# Open and run all cells in train\_diabetes\_model.ipynb  
  
Step 3: Run Flask app  
---------------------  
python app.py  
# Go to http://127.0.0.1:5000 in your browser  
  
Step 4: (Optional) Run Streamlit  
--------------------------------  
streamlit run streamlit\_app.py

# 8. Notes

- Make sure diabetes.csv is in the correct location  
- Create model/ folder if it doesn't exist  
- Use `pip install notebook` if Jupyter isn't installed  
- Use `pip freeze > requirements.txt` to regenerate dependencies