🩺 Capstone Project – AI Chronic Kidney Disease (CKD) Prediction Tool

🚀 *Early screening support using Machine Learning*

Dr. Subramani Suresh

✅ Project Overview

Chronic Kidney Disease (CKD) occurs when the kidneys gradually lose their ability to filter blood effectively. Because CKD progresses silently, early detection is critical to prevent complications such as anemia, heart failure, electrolyte imbalance, and kidney failure.

To support early awareness, I developed an **AI-powered CKD Prediction Web Application** capable of predicting the likelihood of CKD based on simple clinical inputs. This project demonstrates the integration of healthcare data, machine learning, web development, and cloud deployment in a real-world environment.

The system provides fast prediction results and can assist patients, healthcare learners, and early screening programs.

🌐 **Live Demo:**
https://ckd-prediction-6dif.onrender.com

https://subramaniai.pythonanywhere.com/

Github location:

- Model: https://github.com/aisubramani/Hope_Ai_Assignment/blob/main/Week5_7_Web_Project/CKD_Project/Model_RF/Model_creation.ipynb

- Django app: https://github.com/aisubramani/Hope_Ai_Assignment/tree/main/Week5_7_Web_Project/CKD_Project/ckd_project

📊 Dataset Summary

The model uses a dataset containing:

- **400 patient records**
- **24 clinical features**

Example medical indicators include: `['age', 'bp', 'sg', 'al', 'su', 'rbc', 'pc', 'pcc', 'ba', 'bgr', 'bu', 'sc', 'sod', 'pot', 'hrmo', 'pcv', 'wc', 'rc', 'htn', 'dm', 'cad','appet', 'pe', 'ane', 'classification']`

These features are known clinical markers used in CKD diagnosis.
To improve efficiency, feature selection was applied to identify the **6 most influential predictors** using SelectKBest.

🔍 Key Predictive Features

The final model relies on:

✅ Albumin level (urine)
✅ Blood glucose random
✅ Blood urea
✅ Serum creatinine
✅ Packed cell volume
✅ White blood cell count

These values strongly correlate with kidney filtration performance, toxicity levels, and blood concentration.

🧠 Model & Training

After comparative testing, a **Random Forest Classification model** was selected because it provides:

- Strong pattern recognition
- Lower overfitting risk
- Stability with noisy medical data

Training was performed on 80% of the dataset, with 20% reserved for testing.

📈 Model Performance

Metric	Result
Test Accuracy	**97.5%**
ROC-AUC Score	**99.72%**

The ROC-AUC score reflects exceptional classification ability, making the model reliable for early screening.

🧪 Preprocessing Pipeline

To ensure high-quality predictions, the system includes:

- Handling missing values
- Label encoding categorical data

- Standardizing numeric values
- Outlier filtering
- Proper data splitting

All transformations are consistently applied during inference.

🧩 Web Application Integration

The trained model was integrated into a **Django-based web platform**. Users can enter clinical values directly into an interactive form.

The workflow:

1. User submits medical values
2. System loads the trained model
3. Input is preprocessed
4. Prediction result is displayed instantly

The prediction output highlights whether CKD is likely or unlikely.

🎨 User Interface Highlights

The interface includes:

- Clear input guidance
- Highlighted result areas
- Warning banners for positive prediction
- Responsive styling for readability

It is accessible to general users with no technical background.

🎯 Purpose & Practicality

This tool supports:

✅ Early awareness

✅ Educational demonstrations

✅ Screening assistance in remote clinics

✅ Clinical data interpretation practice

While not a medical diagnosis, it can guide timely medical consultation.

🧠 Tech Stack

- Python
- Django
- Scikit-learn
- Pandas
- NumPy

- HTML / CSS

Model serialization is done using Pickle for production use.

🌐 Deployment

The project is hosted on a cloud platform using:

- Render and
-

This ensures performance and easy access anywhere.