**Health Prediction System Documentation**

**Overview**

This document explains our two-stage AI health prediction system that analyzes both monthly clinical data and daily vitals to assess risks for four key health conditions: diabetes, heart disease, sleep apnea, and hypertension.

System Architecture

Our solution consists of two connected AI models:

1. **Monthly Clinical Model**: Uses laboratory test results to predict health risks
2. **Daily Vitals Model**: Incorporates wearable device data and monthly predictions for real-time risk assessment

**Step-by-Step Process**

**Stage 1: Monthly Clinical Predictions**

1. Data Preparation

* **Source**: monthly\_clinical\_data.csv
* **Features**: 18 clinical measurements including:
  + Blood tests (glucose, cholesterol, vitamin levels)
  + Complete blood count (hemoglobin, RBC, WBC)
  + Thyroid function (TSH, T3, T4)
  + Lifestyle factors (smoking, family history)

2. Model Training (training1.py)

* **Preprocessing**:
  + Handles missing values with mean imputation
  + Standardizes all features to equal scale
* **Algorithm**: Random Forest Regressor (one for each condition)
* **Output**: Four trained models saved as .pkl files

3. Prediction API (main1.py)

* **Input**: Clinical test results (17 numeric values + 1 binary)
* **Processing**:
  + Same preprocessing as training
  + Runs all four models
* **Output**: Probability scores (0-1) for each condition

**Stage 2: Daily Vitals Predictions**

1. Data Preparation

* **Sources**:
  + daily\_vitals\_data.csv (wearable device metrics)
  + Monthly predictions from Stage 1
* **Features**: 13 total including:
  + Real-time vitals (pulse, blood pressure, SpO2)
  + Activity metrics (steps, calories)
  + Stress indicators (HRV, stress level)
  + Monthly prediction scores

2. Model Training (training2.py)

* **Preprocessing**:
  + Merges daily and monthly data
  + Standardizes all features
* **Algorithm**: Neural Network (3-layer architecture)
* **Training**:
  + 80/20 train-validation split
  + 50 epochs with early stopping
  + Mean Squared Error loss
* **Output**: Trained model weights (.pth file)

3. Prediction API (main2.py)

* **Input**: 13 daily metrics including 4 monthly predictions
* **Processing**:
  + Applies same scaling as training
  + Neural network inference
* **Output**: Updated probability scores for each condition

**Business Value Proposition**

**Key Benefits**

1. **Comprehensive Risk Assessment**: Combines lab tests with daily monitoring
2. **Early Warning System**: Detects risk changes between clinical visits
3. **Personalized Insights**: Adapts to individual health patterns
4. **Actionable Data**: Clear probability scores for clinical decision support

**Implementation Requirements**

1. **Data Collection**:
   * Monthly clinical tests (standard lab panels)
   * Wearable device data (Fitbit/Apple Watch equivalent)
2. **Infrastructure**:
   * Server to host prediction APIs
   * Secure data storage
3. **Integration**:
   * EHR system connection for clinical data
   * Wearable API connections

**Expected Outcomes**

* 30-50% earlier detection of developing conditions
* 20-40% improvement in preventive intervention targeting
* Continuous risk monitoring between clinical visits
* Personalized health trend analysis

**Maintenance & Improvement**

1. **Model Retraining**: Quarterly updates with new data
2. **Feature Expansion**: Incorporate additional data sources
3. **Validation**: Regular clinical outcome correlation studies

This system provides a powerful bridge between periodic clinical assessments and continuous health monitoring, delivering actionable insights to improve patient outcomes.