Hackathon Check POINT

Goal: Build a product-ready AI/ML prototype that detects early ("weak") health risk signals from longitudinal, self-reported data.

---

1. Context-Aware Detection System

Focus on **changes within a person over time**, not population averages.

Key Approaches

- Compare users with **their own historical baseline**
- Detect **subtle long-term trends**
- Incorporate **life context**

Important Features

- Life events:
    - Job loss
    - Retirement
    - Divorce
    - Major stressors
- Weak signals:
    - Gradual increase in depression score
    - Slow decline in activity or health rating

---

2. Model Performance (60% Weight)

Primary Metric

- **F2-Score (most important)**
    - Emphasizes **recall over precision**

- ○ Missing a sick person = worse than false alarm

Handling Imbalanced Data

- Use:
  - ○ **PR-AUC**
  - ○ Class weighting or resampling

---

3. Practicality and Ethics (40% Weight)

The system must be **real-world usable**, not just research.

Requirements

- **Zero Data Leakage**
  - ○ Do not use features that directly reveal disease
  - ○ Example: medication for the predicted condition
- **Explainability**
  - ○ Show:
    - ■ Why the user was flagged
    - ■ Which features contributed most
- **Fairness**
  - ○ Check bias across:
    - ■ Age
    - ■ Gender
    - ■ Ethnicity
- **Simplicity**
  - ○ Prefer:
    - ■ Lightweight models
    - ■ Fast inference
    - ■ Minimal dependencies

---

4. Required Prototype Functions

A complete submission must include:

1. Data Processing

- Handle:
    - Missing values
    - Noisy self-reported data
    - Longitudinal alignment

2. Risk Scoring

- Output numeric score:
    - Range: **0–100**

3. Risk Categorization

Examples:

- Cardiovascular risk
- Metabolic risk
- Psycho-emotional risk

4. Empathetic Follow-Up

System generates supportive questions:

Example:

> "We noticed some changes in your mood ratings recently. Would you like to share if anything stressful has been happening?"

Rules:

- No diagnosis
- No treatment advice
- Supportive tone

---

5. Strategic Bonus Points

Feature Engineering

- Find **non-obvious correlations**
- Temporal trends
- Behavioral patterns

Production Readiness

- Clean code
- Documentation
- Organized GitHub repository
- 80%+ open-source implementation

Dataset Selection Strategy

Choose dataset based on goal:

- **HRS** → Chronic disease, seniors
- **NLSY97** → Youth transitions
- **PSID** → Health economics and families

Simple Winning Formula

If you remember only this:

Baseline change detection + High recall + Explainability + Clean prototype = Strong submission

---

If you want, I can now help you design a **simple architecture diagram for this hackathon system** (very useful for judges and presentations).