1. Overview

The Health Score Engine is used to assess an individual's health risk based on clinical, lifestyle and demographic data. The health score would be a normalized health score, which is a crucial input in the insurance underwriting process to evaluate risk and determine premiums or eligibility.

---

2. Objectives

- Quantify the applicant's health status as a numeric score (0–100).
- Enable risk-based decision-making during underwriting.
- Standardize health assessments across different data sources.
- Allow explanation or traceability of the score (transparency for underwriters).

---

3. Inputs

We will have input from proposal form and Diagnostic or proposal form, Medical Representative Questions and Diagnostic Reports:

1. Inputs from proposal form.
   Example : Age, Gender, BMI, Occupation, Marital Status, Lifestyle conditions, medications, questions about existing health conditions.

   Following is the link to the Proposal Form Questions:
   ProposalFormQuestions.jpg

2. Following is the link to the questions asked by the Medical Representative
   MERT.pdf

3. Following Diagnostic Reports are collected from the some of the Customers:
   Stress/ Treadmill Test – TMT

   a. Electro CardioGram - ECG
   b. Cotinine Test - CT
   c. Body Mass Index – BMI
   d. Fasting Blood Sugar (FBS)
   e. 2 D Echo
   f. Routine Urine Analysis (RUA)

g. Liver Function Test – LFT
h. Ultrasonography – US (Women Only)
i. Lipid Profile
j. Human Immunodeficiency Virus Test – HIV
k. Chest X-Ray – XR
l. Comprehensive Trail Making Test – CTT

---

4. Outputs

Output Field	Description
Health Score	A score between 0 and 100 indicating overall health (higher is healthier)
Risk Category	e.g., Low / Medium / High
Score Factors	Top 3 contributing features to the score (for explainability)
Underwriting Flag	Pass / Manual Review / Decline

---

5. Processing & Logic

5.1 Preprocessing

- Data normalization & cleaning
- Imputation of missing fields (if feasible)
- Standardization across sources (e.g., units, formats)

5.2 Scoring Algorithm

- Start with a Rule-based model and move to ML model as the data gets accumulated.

6. Functional Requirements

ID	Requirement

FR1	Ingest applicant data from underwriting portal/API
FR2	Validate required fields and flag missing critical data
FR3	Generate health score using configured logic
FR4	Classify score into risk buckets
FR5	Log scoring decision and input snapshot for audit

---

7. Non-Functional Requirements

ID	Requirement
NFR1	Response time < 1s for single applicant scoring
NFR2	Scalable to support bulk scoring (e.g., 10,000 applicants/day)
NFR3	Model retraining support every 3-6 months
NFR4	Secure handling of sensitive health data (HIPAA / IRDAI compliant)
NFR5	High availability with failover strategy

---

8. Integrations

- **Underwriting Portal / CRM**
- **Policy Admin System** (optional)

---

9. Assumptions

- Minimal dataset size available to train initial model.
- Clinical text/NLP support is optional and phase-2.
- Health data is structured/formatted before ingestion.

---

10. Future Enhancements

- Integration (ABDM) Ayushman Bharat Digital Mission - ABHA.
- Risk Assessment ML models
- Personalized premium recommendations.

Simple Rule Engine Example

What to Solve:

Simple weighted scoring approach combining **lifestyle** (behavioral) inputs and **CBC** (Complete Blood Count) lab values into a single **0–100 health score**.

Design overview (high level)

- Output: single **Health Score (0–100)**.
- Components:
    - **Lifestyle** (weight = 40% of total score → 0–40 points)
    - **CBC** (weight = 60% of total score → 0–60 points)
- Each component is the sum of sub-scores computed by simple rules (thresholds, boolean checks).
- Final score = lifestyle_score + cbc_score (rounded).

Lifestyle (0–40 points total)

Weights inside lifestyle:

- Smoking: 0–10 points
    - Non-smoker → 10
    - Former smoker (quit >1 year) → 7
    - Occasional smoker → 4
    - Regular smoker → 0
- Alcohol consumption: 0–6 points
    - None/rare → 6

    - Moderate (within guidelines) → 4
    - Above guidelines → 1
- Physical activity: 0–10 points
    - ≥150 min/week moderate or ≥75 min vigorous → 10
    - 75–149 min moderate → 6
    - <75 min → 2
- Diet quality: 0–8 points (subjective / survey)
    - Healthy balanced diet → 8
    - Mixed → 4
    - Poor → 0
- Sleep: 0–6 points
    - 7–9 hours good quality → 6
    - 6–7 or 9–10 → 3
    - <6 or poor quality → 0

(These add up to 40.)

CBC (0–60 points total)

Pick common CBC markers — each mapped to 0–12 points so 5 markers * 12 = 60 (you can choose different splits).

Example markers and *example* reference ranges (adapt to lab & patient sex/age):

- Hemoglobin (Hb) — good range:
    - Female: 12–16 g/dL, Male: 13.5–17.5 g/dL
    - Scoring (0–12): full points if within range, partial if mildly low/high, 0 if severe abnormal.
- White Blood Cell (WBC) count — good range: 4.0–11.0 ×10^9/L
- Platelet count — good range: 150–400 ×10^9/L
- Mean Corpuscular Volume (MCV) — good range: 80–100 fL
- RBC morphology/flags (or neutrophil% / lymphocyte% if you prefer) — objective check or substitute another numeric CBC measure.

Scoring approach per marker (0–12):

- Within normal range → 12
- Mildly abnormal (10–20% outside range) → 6–8
- Moderately abnormal (20–40% outside) → 2–4
- Severely abnormal (>40% outside) → 0

You can also penalize combinations (e.g., both low Hb and low MCV → iron deficiency pattern → lower points).

Aggregation & interpretation

- Final numeric score = lifestyle_points (0–40) + cbc_points (0–60) → round to nearest integer.
- Interpretation example:
    - 85–100 → Excellent
    - 70–84 → Good
    - 50–69 → Fair — consider lifestyle changes / checkups
    - <50 → Concerning — recommend clinical follow-up