Problem Statement Document

1. Introduction

This hackathon invites participants to design and implement innovative Artificial Intelligence and Machine Learning solutions that address high-impact, real-world challenges. Competitors will work with structured, unstructured, or multimodal datasets to build models that are accurate, interpretable, and production-ready. Each track includes unique technical questions to push creativity and rigorous thinking.

2. Problem Tracks

Track 1: Predictive Healthcare Risk Modeling

Participants will receive anonymized patient datasets with demographics, vitals, historical diagnostics, lifestyle indicators, and time-series lab results. The task is to predict the likelihood of early onset of a chronic disease within the next 12 months.

Key Questions:

1. How will you design a pipeline that handles missing, sparse, and noisy clinical data without data leakage?
2. Can you create an interpretable model that highlights risk-driving features while ensuring medical plausibility?
3. How will you test and mitigate demographic bias in your model's predictions?
4. What metrics will you use beyond accuracy to ensure clinical reliability?

Track 2: Real-Time Retail Demand Forecasting

You are provided multi-store retail sales data, including price, promotions, stock availability, competitor activity, and seasonal patterns. Build a forecasting model for daily demand for the next 14 days.

Key Questions:

1. What architectural choice best balances long-term recurrence (seasonality) and short-term fluctuations (promotions)?
2. How do you ensure the model generalizes across stores with different customer profiles?
3. How will you detect and adapt to concept drift resulting from changing market behavior?
4. Can you design an incremental learning or streaming-based forecasting module?

Track 3: Toxicity Detection and AI-Assisted Moderation

Build an NLP model capable of identifying harassment, hate speech, implicit hate, coded language, and contextual toxicity in text messages or social media comments.

Key Questions:

1. How will you ensure the model differentiates genuine toxicity from sarcasm, reclaimed slurs, or sensitive cultural expressions?
2. How will you reduce false positives that unnecessarily censor users?
3. Can you design a lightweight model capable of running at real-time latencies?
4. Which methods will you use for explainability to justify moderation outcomes?

Track 4: Computer Vision for Environmental Risk Detection

You will receive drone imagery of forests, water bodies, and agricultural land. Build a vision system that identifies hazards such as illegal deforestation, industrial waste discharge, or crop disease.

Key Questions:

1. How will you deal with severe class imbalance, where hazardous events occur infrequently?
2. What image augmentation and preprocessing steps will improve robustness under low-light or noisy imaging conditions?
3. Can you integrate spatial metadata (GPS/altitude) into the model pipeline?

3. Deliverables

Participants must submit the following:

1. **Source Code Repository**
    - Clean, modular, reproducible code.
    - Setup instructions and environment files.

2. **Model Report**
    - Problem interpretation, approach, and reasoning.
    - Data preprocessing steps and EDA summaries.
    - Model design, evaluation metrics, and error analysis.
    - Explainability and fairness evaluation.

3. **Final Presentation (PDF or PPT)**
    - 5–10 slides summarizing your solution and results.

4. **Optional Demo**
    - Live demo or short recorded video demonstrating usage.

4. Evaluation Criteria

1. **Innovation & Originality**
   Novelty in model design, feature engineering, or problem framing.
2. **Model Performance**
   Accuracy, robustness, and generalization on both public and hidden datasets.
3. **Scalability & Efficiency**
   Ability of the solution to handle larger datasets and real-time contexts.
4. **Explainability & Fairness**
   Interpretability and mitigation of bias across subgroups.
5. **Documentation & Presentation Quality**
   Clarity, completeness, and professional structure.

5. Rules & Guidelines

- Teams may consist of 1–3 members.
- Use of open-source libraries and pretrained models is allowed unless specified.
- All solutions must be original and created during the hackathon period.
- Plagiarism or unethical data usage results in immediate disqualification.

We look forward to innovative, impactful, and responsible AI/ML solutions from all participants.