Capstone Project Concept Note and Implementation Plan

Project Title: AI for Maternal Health Risk Prediction

1. Project Overview

This project addresses the high maternal mortality rate in low-resource settings due to a lack of early risk detection during pregnancy. Our goal is to develop an AI-powered web/mobile platform that uses a machine learning classification model to predict the risk level of pregnancy complications. By providing a simple, questionnaire-based interface, the system will assist expectant mothers and healthcare workers in identifying high-risk pregnancies early and seeking appropriate care. This project aligns with SDG 3 (Good Health), SDG 5 (Gender Equality), and SDG 10 (Reduced Inequality).

2. Objectives

- Develop a machine learning model to classify maternal health risks (Low, Medium, High).
- Integrate psychosocial indicators (e.g., stress, sleep, nutrition) into prediction.
- Design and deploy a mobile/web-based interface for real-world usage
- Collaborate with local clinics or NGOs to refine and validate the system.
- Raise awareness and empower women in low-resource communities to take early action

3. Background

Maternal mortality remains a global health challenge, particularly in regions with limited access to medical care and diagnostic tools. Existing solutions rely primarily on physiological data and clinical intervention. However, many women in underserved areas do not receive routine medical screenings. This project introduces a machine learning approach that incorporates both clinical and psychosocial indicators, providing a more accessible and holistic tool for early risk assessment.

4. Methodology

We will employ supervised machine learning for multi-class classification of maternal health risks. Models such as Random Forest, Logistic Regression, and XGBoost will be trained using publicly available and augmented datasets. Feature engineering will include standard physiological attributes and additional simulated psychosocial factors. Performance will be evaluated using accuracy, precision, recall, F1-score, and model interpretability via SHAP or LIME.

5. Architecture Design Diagram

The system architecture consists of the following components:

- Data Input: User questionnaire via web/mobile interface
- Preprocessing Module: Cleans, encodes, and scales input data.
- ML Model: Predicts maternal risk level
- Explainability Layer: SHAP/LIME for feature attribution
- Output Interface: Displays risk level and recommendations

6. Data Sources

We will use the UCI Maternal Health Risk dataset (CSV, ~1,000 records) as our primary dataset. Additional data from India’s RMNCH+A reports and WHO regional case studies will be incorporated to improve generalizability. Simulated features such as stress level and nutrition scores will be used to represent psychosocial factors until real-world data is obtained.

7. Literature Review

Studies by Smith et al. (2021) and Kumar & Patel (2022) demonstrate the viability of ML in maternal risk prediction, focusing on physiological features. Our project extends this work by incorporating psychosocial indicators and targeting low-resource settings through a lightweight, interpretable model with local deployment potential.

Implementation Plan

1. Technology Stack

- Programming Languages: Python, JavaScript
- ML Libraries: scikit-learn, XGBoost, SHAP
- Web Frameworks: Flask or FastAPI (backend), React (frontend)
- Data Handling: Pandas, NumPy, CSV files
- Deployment: Heroku or local server (prototype phase)

2. Timeline

- Week 1-2: Literature review, feature planning, data exploration
- Week 3-4: Model training and evaluation
- Week 5: App interface design (frontend/backend)
- Week 6: Integration and deployment
- Week 7: Usability testing with peers or NGO partners
- Week 8: Final documentation and presentation

3. Milestones

- Dataset ready and features engineered.
- Model achieves ≥85% accuracy with acceptable F1-score
- Interface receives input and predicts risk successfully.
- SHAP plots integrated for transparency.
- Final system deployed and tested

4. Challenges and Mitigations

- Data limitations: Use augmentation and simulation techniques
- Model overfitting: Use cross-validation and regularization
- UX design for low-literacy users: Test with real users and simplify interface
- Internet access constraints: Enable offline mode in future iterations

5. Ethical Considerations

All data used will be anonymized and sourced from public repositories or simulated. The system will clearly state that it is not a diagnostic tool but a decision support aid. We will monitor for bias in prediction outcomes and apply fairness checks during evaluation.