Project Proposal: Heart Disease Risk Classification with Time-Series Analysis

Project Title: Heart Disease Risk Classification Using Temporal Clinical Patterns

1. Problem Statement
Cardiovascular disease is a leading global health concern. While machine learning models can predict heart disease risk using clinical measurements from a single point in time, this ignores potentially valuable information from how a patient's vitals change over time. This project will explore whether adding **time-series analysis** of key clinical measurements improves prediction accuracy compared to using only the most recent data. We will use the standard UCI Heart Disease dataset enhanced with simple synthetic time-series data to test this hypothesis.

2. Dataset & Approach

- **Primary Data:** Heart Disease UCI dataset from Kaggle (14 clinical features, binary classification target)

- **Time-Series Enhancement:** We will create simple simulated historical data for 3 key features that naturally change over time: resting blood pressure, maximum heart rate, and cholesterol. Each patient will have 6 monthly historical values generated using basic trend simulation.

- **Rationale:** This approach lets us apply time-series techniques while keeping the project manageable and interpretable.

3. Specialization Focus: Time Series Analysis
This project will incorporate **time-series feature engineering** as the specialization component. Instead of just using the latest measurements, we will:

1. Extract trends from the simulated historical data

2. Create features like rolling averages, slopes, and variability measures

3. Compare model performance with and without these temporal features

4. Proposed Methodology

Phase 1: Data Preparation & Exploration

- Load and explore the standard UCI dataset
- Generate simple time-series data for key vitals using basic statistical simulation
- Visualize both static data distributions and example patient trends

Phase 2: Feature Engineering

- Process static features: handle missing values, encode categories, scale
- From time-series data, create summary features:
    - Trend direction (improving/worsening/stable)
    - Average values over last 3 months
    - Variability measures
- Combine static and temporal features into final dataset

Phase 3: Modeling & Comparison

- **Baseline:** Train models using only original static features
- **Enhanced:** Train same models using static + temporal features
- Use simple algorithms: Logistic Regression and Random Forest
- Evaluate both approaches using accuracy, precision, recall, and AUC-ROC

Phase 4: Analysis & Reporting

- Compare performance between baseline and enhanced models
- Identify which temporal features contribute most to predictions
- Document findings on whether temporal patterns add predictive value

5. Project Scope & Limitations

- Time-series data will be simulated (not real historical data)
- Focus is on demonstrating the time-series approach rather than production-level accuracy
- Simple, interpretable models will be prioritized over complex ensembles

Conclusion

This project will apply time-series analysis techniques to a classic classification problem, demonstrating how temporal patterns in clinical data might improve heart disease risk prediction. The approach keeps the core classification task straightforward while adding a clear specialization component through time-series feature engineering.