

1. Problem Scope

Problem Statement

Hospitals often face high patient readmission rates, which increase healthcare costs and reduce care quality. The goal is to build an AI model that predicts whether a patient is at risk of being readmitted within 30 days of discharge.

Objectives

Identify high-risk patients before discharge.

Enable proactive interventions (e.g., follow-ups, medication tracking).

Reduce readmission rates and penalties under value-based care models.

Stakeholders

Hospital Management: Interested in reducing costs and improving quality metrics.

Doctors and Nurses: Use the predictions to tailor post-discharge care.

Patients: Receive timely care and better outcomes.

2. Data Strategy

Data Sources

Electronic Health Records (EHR): Past medical history, medications, comorbidities, previous admissions.

Demographic Data: Age, gender, income, insurance type, address.

Discharge Summary & Notes: Length of stay, vitals, lab results at discharge.

Ethical Concerns

Patient Privacy & HIPAA Compliance: All patient data must be encrypted and anonymized to comply with regulations.

Bias in Healthcare Data: Socioeconomic and racial bias could lead to discriminatory predictions (e.g., patients from certain zip codes flagged unfairly).

Preprocessing Pipeline

Handle Missing Values: Use imputation for missing vitals or notes.

Feature Engineering:

Calculate Charlson Comorbidity Index as a single severity feature.

Create a readmission count in past 6 months feature.

Extract discharge weekday (weekend vs weekday).

Encode Categorical Variables: Insurance type, gender, etc. via one-hot encoding.

Normalize Numerical Features: Age, length of stay, etc.

3. Model Development

Model Selection: Gradient Boosting (e.g., XGBoost)

Justification: Works well with tabular data, handles missing values, and captures complex interactions between medical features.

Confusion Matrix (Hypothetical)

Predicted: Yes (0)	Predicted: No (1)
Actual: Yes (0)	80 20
Actual: No(1)	25 75

Precision & Recall:

Precision = $TP / (TP + FP) = 80 / (80 + 25) = 0.76$

Recall = $TP / (TP + FN) = 80 / (80 + 20) = 0.80$

4. Deployment

Integration Steps:

Containerize the model (e.g., Docker).

Use a REST API (Flask/FastAPI) to connect the model to the hospital's internal EHR system.

Trigger predictions at discharge time based on patient data.

Show output in the clinician dashboard with recommended action flags.

Compliance with Regulations:

Use end-to-end encryption for all patient data.

Apply role-based access control (RBAC).

Conduct regular audits and use de-identified training data.

Store predictions and decisions in compliance with HIPAA & GDPR.

5. Optimization

Preventing Overfitting:

Use cross-validation + early stopping during training.

Cross-validation ensures the model generalizes well.

Early stopping halts training when validation error stops improving.