Part 2: Case Study Application (40 points)

Problem Scope (5 points)

Problem Statement:
Develop an AI system to predict the likelihood of patient readmission within 30 days after hospital discharge.

Objectives:

- Reduce avoidable hospital readmissions by 15%.
- Improve personalized post-discharge care planning.
- Optimize hospital staffing and resource allocation.

Stakeholders:

- Hospital administrators, medical directors, and clinicians.
- Patients and caregivers who rely on accurate follow-up care.

Data Strategy (10 points)

Proposed Data Sources:

- **Electronic Health Records (EHRs):** Clinical notes, diagnosis history, lab test results.
- **Patient Demographics:** Age, gender, ethnicity, socioeconomic status.
- **Admission & Discharge Records:** Length of stay, primary diagnosis, discharge summary.

Ethical Concerns:

1. **Patient Privacy:** Ensure all data is anonymized and managed under HIPAA-compliant protocols.
2. **Bias & Fairness:** Avoid algorithmic bias that could marginalize underrepresented or vulnerable populations.

Preprocessing Pipeline:

- **Data Cleaning:** Remove duplicates, fix incorrect values, and address missing fields.
- **Feature Engineering:**
    - Days since last admission
    - Comorbidity scores (chronic conditions count)
    - Total number of prior readmissions

- **Encoding & Normalization:**
    - One-hot encode categorical features (e.g., gender, insurance type)
    - Normalize continuous variables (e.g., age, lab results)

Model Development (10 points)

Selected Model:
Gradient Boosting Machine (XGBoost) – chosen for its superior performance on structured data, tolerance to missing values, and model interpretability through feature importance.

Hypothetical Confusion Matrix:

	Predicted Readmit	**Predicted No Readmit**
Actual Readmit	80	20
Actual No Readmit	15	85

Evaluation Metrics:

- **Precision:** 8080+15=0.842\frac {80} {80 + 15} = 0.84280+1580=0.842
- **Recall:** 8080+20=0.800\frac {80} {80 + 20} = 0.80080+2080=0.800

These scores indicate the model accurately identifies high-risk patients while minimizing false positives.

Deployment (10 points)

Integration Plan:

- **Containerization:** Use Docker to encapsulate the model and its environment.
- **Model Serving:** Build an API endpoint with Flask or FastAPI.
- **EHR Integration:** Connect the model to hospital systems via RESTful APIs for real-time predictions at discharge.

Compliance Measures:

- Implement full encryption of data in transit and at rest.
- Enforce Role-Based Access Control (RBAC) to limit sensitive data access.
- Maintain detailed audit logs and perform regular HIPAA compliance audits.

Optimization (5 points)

Overfitting Prevention Strategy:

- Apply **k-fold cross-validation** during training.
- Use **L1/L2 regularization** to prevent over-complex models.
- Enable **early stopping** to halt training when validation loss stops improving.