🚑 Part 2: Case Study – Hospital Readmission Prediction (40 points)

1. Problem Scope (5 points)

Problem:
Hospitals aim to predict whether a patient will be readmitted within 30 days of discharge. This enables proactive interventions, improves care quality, and reduces operational costs.

Objectives:

1. Predict patient readmission risk using AI models
2. Support discharge planning decisions for clinicians
3. Reduce unnecessary hospital readmissions

Stakeholders:

- Doctors & hospital staff
- Patients and their families
- Hospital administrators & insurance providers

2. Data Strategy (10 points)

Data Sources:

- **Electronic Health Records (EHRs):** Medical history, vitals, medications, discharge summaries
- **Demographic Data:** Age, gender, location, lifestyle, insurance info

Ethical Concerns:

1. **Patient Privacy:** Mishandling of sensitive medical data could violate patient trust and regulations like HIPAA
2. **Algorithmic Bias:** Models may unfairly predict higher risk for marginalized groups due to historical bias in data

Preprocessing Pipeline:

Step	Description
1. **Data Cleaning**	Handle missing values in records (e.g., BMI, diagnosis)
2. **Feature Engineering**	Derive features like "# of past admissions", "comorbidity score", "length of stay"
3. **Encoding & Scaling**	One-hot encode categorical variables (e.g., gender), normalize numerical features like age or lab results

3. Model Development (10 points)

Chosen Model:
🧠 **Random Forest Classifier**

Why:

- Works well with tabular EHR data
- Handles non-linear relationships
- Provides feature importance insights (great for healthcare explainability)

Hypothetical Confusion Matrix (Test Set, 100 patients):

	Predicted: No Readmission	Predicted: Readmission
Actual: No	60	10
Actual: Yes	5	25

Metrics:

- **Precision:** 25 / (25 + 10) = **0.714**
- **Recall:** 25 / (25 + 5) = **0.833**

🎯 High recall ensures we catch most at-risk patients. Precision ensures we don't flood doctors with false alarms.

---

4. Deployment (10 points)

Deployment Steps:

1. Integrate with hospital's EHR system via secure APIs
2. Host model on an internal cloud or local server
3. Display readmission risk scores in doctors' dashboards
4. Trigger alerts when high-risk patients are being discharged
5. Log predictions for auditing and retraining

Compliance Strategy (e.g., HIPAA):

- Encrypt all patient data (at rest and in transit)
- Use access control + audit logs
- Get patient consent for data usage if required

---

5. Optimization (5 points)

Preventing Overfitting:
☑ **Cross-validation + Feature Selection**

- Use k-fold cross-validation to generalize performance
- Drop irrelevant features (like ID fields or duplicated metrics)
- Apply regularization if switching to logistic regression or neural nets