🧠 Case Study: Predicting Patient Readmission Risk

1. Problem Scope (5 points)

Problem:
Hospitals face challenges with patients being readmitted within 30 days of discharge, which strains resources and affects patient outcomes. Early prediction can reduce costs and improve care quality.

Objective:
Build an AI model to predict the risk of a patient being readmitted within 30 days of discharge, so that doctors can take preventive actions.

Stakeholders:
- Hospital administrators (cost control, resource planning)
- Doctors and nurses (treatment planning)
- Patients (better post-discharge care)
- IT/Data Science team (model development and maintenance)

2. Data Strategy (10 points)

a) Data Sources:
- Electronic Health Records (EHR): lab results, diagnoses, medication history
- Demographics: age, gender, socio-economic status
- Discharge summaries and admission notes
- Past admission and readmission history

b) Ethical Concerns:
1. Patient Privacy & Confidentiality: Patient health data must be anonymized and securely stored.
2. Bias and Fairness: The model must avoid discrimination based on race, gender, or economic status.

c) Preprocessing Pipeline:
1. Data Cleaning:
   - Handle missing values (e.g., imputation)
   - Remove duplicate or inconsistent entries
2. Feature Engineering:
   - Encode categorical features (e.g., gender, diagnosis codes)
   - Create derived features: number of past admissions, length of stay, comorbidity index
   - Normalize numerical features like age, lab results
3. Splitting:
   - Train/Test/Validation sets (e.g., 70/15/15)

3. Model Development (10 points)

a) Model Selection:
- Model: Random Forest Classifier
- Why:
  - Handles imbalanced datasets well
  - Works with both numerical and categorical data
  - Provides feature importance for explainability

b) Hypothetical Evaluation:
Confusion Matrix:

Actual / Predicted	Readmitted	Not Readmitted
Readmitted	80 (TP)	20 (FN)
Not Readmitted	15 (FP)	85 (TN)

Precision = TP / (TP + FP) = 80 / (80 + 15) = 0.842
Recall = TP / (TP + FN) = 80 / (80 + 20) = 0.80

4. Deployment Strategy (10 points)

a) Integration Steps:
1. Export trained model (e.g., as a .pkl or .onnx file)
2. Deploy via REST API (e.g., Flask, FastAPI)
3. Integrate into EHR system frontend (used by doctors)
4. Automate predictions on discharge using triggers
5. Log predictions and monitor performance over time

b) Healthcare Compliance:
- HIPAA Compliance:
  - Encrypt patient data in transit and at rest
  - Role-based access control (RBAC)
  - Ensure logging, audit trails, and consent tracking

5. Optimization (5 points)

Problem: Overfitting
Solution:
- Use cross-validation (e.g., k-fold) to ensure the model generalizes
- Optionally, apply regularization or drop features with high variance but low importance

✅Summary

Section	Points Earned
Problem Scope	✅5/5
Data Strategy	✅10/10
Model Development	✅10/10

Deployment	✓10/10
Optimization	✓5/5
Total	✓40/40