AI Assignment Explanation: Patient Readmission Risk Prediction

Part 1: Short Answer Questions

1. Problem Definition

Hypothetical AI problem: Predicting student dropout risk within one academic year.

Objectives:

1. Identify students at high risk of dropping out before the end of the academic year.
2. Provide actionable risk scores and top contributing factors per student.
3. Reduce overall dropout rate by 10% within one year of model deployment.

Stakeholders: School administrators, students and families.
KPI: Recall for dropout class to ensure high-risk students are identified.

2. Data Collection & Preprocessing

Data sources: LMS and attendance records, administrative and demographic data.
Potential bias: Sampling bias due to missing data from students who already left school.
Preprocessing steps include handling missing data, normalization, and feature encoding.

3. Model Development

Chosen Model: Random Forest because it handles mixed data types and is interpretable.
Data split: 70% training, 15% validation, 15% test.
Hyperparameters tuned: n_estimators and max_depth to control complexity and performance.

4. Evaluation & Deployment

Metrics: Precision and Recall for identifying true high-risk students.
Concept Drift: Occurs when the relationship between inputs and targets changes over time. Monitor using PSI and periodic re-evaluation.
Deployment Challenge: Scalability of inference for large student populations.

Part 2: Case Study — Hospital Readmission Prediction

Objective: Predict patients' 30-day readmission risk after discharge using EHR data.
Stakeholders: Clinicians, patients, and hospital administrators.

Data Strategy:

1. Data Sources: EHRs, demographic and utilization history.
2. Ethical Concerns: Patient privacy and algorithmic fairness.
3. Preprocessing: Data linkage, imputation, aggregation, feature engineering, and scaling.

Model Development:

Model: Gradient Boosted Trees (LightGBM/XGBoost).
Evaluation: Confusion matrix generated with 80 TP, 30 FN, 20 FP, and 370 TN. Precision = 0.8, Recall = 0.73.

Deployment Steps:

1. Integrate model predictions into hospital discharge workflows.
2. Ensure compliance with HIPAA and local privacy laws.
3. Set up monitoring for model performance and drift.

Optimization: Use regularization and early stopping to address overfitting.

Part 3: Critical Thinking

Ethics & Bias:

Biased data can lead to unfair risk predictions, affecting care quality. Strategy: subgroup performance audits and fairness-aware reweighting.

Trade-offs:

Balance between interpretability and accuracy is crucial. Hospitals often prefer interpretable models even if slightly less accurate.
Limited computational resources may push teams to choose lightweight models such as logistic regression.

Part 4: Reflection

The most challenging step is data cleaning and integration due to missing or inconsistent EHR data.
Future improvements: Acquire richer datasets, automate ETL, and improve clinician feedback loops.