AI in Healthcare – Predicting Patient Readmission Risk

Part 1: Short Answer Questions (30 points)

1. Problem Definition (6 points)

Problem: Predicting patient readmission risk within 30 days after hospital discharge.

Objectives:

1. Identify patients at high risk of readmission.
2. Optimize resource allocation for post-discharge care.
3. Reduce hospital readmission rates and improve patient outcomes.

Stakeholders:

- Hospital administrators
- Healthcare providers (doctors, nurses)

Key Performance Indicator (KPI):

- Area Under the ROC Curve (AUC) to measure prediction accuracy of high-risk readmissions.

2. Data Collection & Preprocessing (8 points)

Data Sources:

1. Electronic Health Records (EHRs) – patient vitals, diagnosis, treatment history.
2. Demographics & Social Determinants – age, gender, insurance, socioeconomic factors.

Potential Bias:

- Historical data may overrepresent certain patient groups, leading to biased predictions against minority populations.

Preprocessing Steps:

1. Handle missing data via imputation or removal of incomplete records.
2. Normalize continuous variables to a common scale.
3. Encode categorical variables using one-hot encoding.

3. Model Development (8 points)

Model Choice: Random Forest Classifier

Data Split: Training set: 70%, Validation set: 15%, Test set: 15%

Hyperparameters to Tune:

1. Number of trees (n_estimators)

2. Maximum tree depth (max_depth)

4. Evaluation & Deployment (8 points)

Evaluation Metrics:

1. Precision
2. Recall

Concept Drift:

- Occurs when patient population or treatment protocols change over time, affecting model accuracy.

Technical Challenge:

- Scalability: Ensuring the model can handle hospital-wide data in real-time.

Part 2: Case Study Application (40 points)

Problem Scope (5 points)

Problem: Predict 30-day readmission risk to reduce unnecessary hospital readmissions.

Objectives:

1. Reduce readmission rates
2. Allocate post-discharge resources efficiently
3. Improve patient outcomes

Stakeholders:

- Hospital administrators
- Patients
- Healthcare providers

Data Strategy (10 points)

Proposed Data Sources:

1. EHRs
2. Patient demographics and insurance data

Ethical Concerns:

1. Patient privacy and data security
2. Potential algorithmic bias

Preprocessing Pipeline:

1. Handle missing values using median imputation
2. Feature engineering: create a “comorbidity score”

3. Normalize numerical features and encode categorical features

Model Development (10 points)

Model Choice: Random Forest Classifier

Confusion Matrix (Hypothetical Data):

	Predicted High	Predicted Low
Actual High	45	5
Actual Low	10	40

- Precision: 0.82
- Recall: 0.90

Deployment (10 points)

Integration Steps:

1. Wrap model in a REST API
2. Connect API to hospital's EHR system
3. Schedule batch predictions daily

Compliance:

- Ensure HIPAA compliance by encrypting data and limiting access

Optimization (5 points):

- Use cross-validation and regularization to reduce overfitting

Part 3: Critical Thinking (20 points)

Ethics & Bias (10 points)

- Biased training data may underestimate readmission risk for minority groups.
- Mitigation: re-sampling or weighting techniques, fairness metrics.

Trade-offs (10 points)

- Interpretability vs. Accuracy: Random Forest is interpretable, deep learning may be more accurate.
- Limited Resources: Prefer simpler models like Random Forest over large neural networks.

Part 4: Reflection & Workflow Diagram (10 points)

Reflection (5 points)

- Challenges: Feature engineering due to diverse patient data

- Improvements: Collect longitudinal data and implement automated bias detection tools

Workflow Diagram (5 points