Week 5 assignment part 2

◆ **1: Problem Scope (5 points)**

Problem

Hospitals face both financial and clinical challenges due to patients being readmitted within 30 days of discharge. Reducing unnecessary readmissions is a key priority for improving healthcare quality and efficiency.

Objective

Develop an AI system that predicts the likelihood of a patient being readmitted within 30 days of discharge, so healthcare staff can take preventive measures.

Stakeholders

- **Hospital Management** – reduce costs, improve quality scores.

- **Clinicians (Doctors & Nurses)** – use risk predictions to guide follow-up care.

- **Patients** – receive proactive and personalized care.

- **Data & IT Teams** – support data integration, privacy, and deployment.

◆ **2: Data Strategy (10 points)**

a) Proposed Data Sources

- **Electronic Health Records (EHR)**: admission/discharge dates, diagnoses (ICD-10), procedures, lab results.

- **Patient Demographics**: age, gender, socioeconomic status.

- **Admission History**: prior readmissions, chronic conditions.

- **Medications**: prescribed drugs at discharge.

b) Ethical Concerns

1. **Patient Privacy**: Medical data is sensitive and must be protected (HIPAA compliance).

2. **Algorithmic Bias**: Data may underrepresent certain populations (e.g., rural or minority patients), leading to unfair predictions.

◆ **3: Model Development (10 points)**

a) Model Selection and Justification

Random Forest Classifier

- Handles mixed data types (numerical & categorical).

- Reduces overfitting through ensemble learning.

- Provides feature importance for interpretability.

◆ **4: Deployment (10 points)**

a) Integration into Hospital Systems

1. **Model Packaging**: Save model using joblib or pickle.

2. **API Deployment**: Use **Flask** or **FastAPI** to expose predictions via REST API.

3. **EHR Integration**: Connect to hospital systems like Epic or Cerner to run predictions at discharge.

4. **Clinician Dashboard**: Show readmission risk score and top contributing features.

5. **Monitoring**: Log predictions and flag drift or unexpected results.

b) Regulatory Compliance

- **Data Encryption**: Secure data at rest and in transit (TLS/HTTPS).

- **Access Control**: Ensure only authorized staff can access predictions.

- **Audit Logging**: Record access to model inputs/outputs.

- **De-identification**: Remove or mask sensitive data used for training.

- **HIPAA Compliance**: Work with legal teams to align with healthcare regulations

◆ **5: Optimization (5 points)**

🛠️ **Method to Address Overfitting:**

Cross-Validation + Regularization

- Use **K-Fold Cross-Validation** to ensure the model generalizes across different subsets of data.

- In tree-based models like Random Forest:

    ○ Limit max_depth

    ○ Increase min_samples_split

    ○ Tune n_estimators