AI Development Workflow- Case Study: Prediction of Patient Readmission

Part 2: Application of Case Study (40 points)

Problem Scope (5 points)

One of the hospitals is interested in introducing an AI that can predict readmission within 30 days following discharge. This is beneficial in cutting down on preventable readmission and enhancing patient care.

Objectives:

i. Forecast readmission chances based on patient records.
ii. Identify potentially at-risk patients to be proactively intervened in.
iii. Enhance hospital performance and minimize cost charges as a result of readmission excess.

Stakeholders:

i. Care managers and hospital administrators.
ii. Families and patients.

Data Strategy (10 pt)

Recommended Data sources:

1. Vitals, Visits, Electronic Health Records (EHRs), and diagnoses.
2. Demographics and outpatient clinic history (simulated by patients.csv).

Ethical Concerns:

i. Patient privacy: Protected health information should not be dealt with carelessly.
ii. Bias: Older data may induce bias (e.g., underdiagnosing marginalized groups).

Preprocessing Pipeline:

- Excluded rows having a missing target (readmit).
- Missing numeric values coded as imputed (e.g., age, bp).
- Categorical features encoded using one-hot encoding (gender, clinic, diagnosis).
- Standard scaler normalized continuous features.
- Train/test split to assess generalization.

Model Development (10 points)

The Model Selected: XGBoostClassifier

❖ Justification: It works well with structured/tabular data, supports missing values, and has high performance and interpretable results.

Output Confusion Matrix (evaluation):

[[6 4]

[1 9]]

Classification Report:

- Accuracy 75 percent
- Precision (class 1): 0.69
- Recall (class 1) = 0.90

Precision/Recall Explanation:

- Good recall guarantees that most at-risk patients will be recruited.
- Accuracy translates to the precision of those predictions.

Deployment (10 points)

Integration Steps:

i. The model and scaler are fitted and stored as .joblib files.
ii. The FastAPI app presents /predict endpoint.
iii. It takes patient data as JSON and outputs the prediction.
iv. Can be tested with Swagger UI or curl.

Healthcare Compliance:

- Even though it is simulated, this will have to follow rules such as HIPAA or the Data Protection Act in real life.
- Has adequate logging, encoding, and consent measures.

Technical Issue Afflicted:

- Feature mismatch: The Prediction API should be identical to the training. Addressed by hard-coding the anticipated features and by predicting missing dummy columns.

Optimization (5 points)

Overfitting Mitigation:

i. Tested generalization with a train-test split.
ii. Could also apply:
    - Regularization (lambda, alpha in XGBoost)
    - Cross-validation
    - Validation loss early stopping