

AI for Software Engineering - Final Assignment Report

Part 1: Short Answer Questions

1. Problem Definition

Hypothetical AI Problem: Predicting student dropout rates in online learning platforms.

Objectives:

1. Identify at-risk students early to provide timely interventions.
2. Improve overall student retention and success rates.
3. Help institutions allocate support resources efficiently.

Stakeholders:

1. Students
2. Educational Administrators

KPI: Percentage reduction in student dropout rates after implementing the AI system.

2. Data Collection & Preprocessing

Data Sources:

1. LMS logs (e.g., login frequency, video watch time)
2. Demographic and academic performance data

Potential Bias: Socioeconomic bias due to incomplete data

Preprocessing Steps:

1. Handle missing values
2. Normalize features
3. Encode categoricals

3. Model Development

Model: Random Forest

Data Split: 70/15/15

Tune: number of trees, max depth

4. Evaluation & Deployment

Metrics: Accuracy, F1-Score

Concept Drift Monitoring

Deployment Challenge: LMS integration

Part 2: Case Study Application

Problem: Predict 30-day readmission post-discharge

Stakeholders: Hospital, Patients

Data: EHR + Demographics

Ethics: Privacy + Bias

Pipeline: Missing values, Normalization, One-hot encoding, Feature engineering

Model: Random Forest

Confusion Matrix:

TP=45, FN=15, FP=10, TN=130

Precision: 0.818, Recall: 0.75

Deployment: FastAPI, Streamlit, HIPAA compliance

Overfitting: K-fold + pruning

Part 3: Critical Thinking

Bias can harm minority patient predictions. Mitigation: fairness-aware algorithms + monitor performance by subgroup.

Interpretability vs Accuracy: simpler models are explainable, deep models more accurate.

Limited resources favor simpler models.

Part 4: Reflection & Workflow Diagram

Reflection: Hardest was designing compliance-friendly, bias-aware, scalable pipeline.

Improvement: Use federated learning + deep learning.

Diagram included separately as PNG.