AI Development Workflow Assignment

Course: AI for Software Engineering

Duration: 7 Days

Total Points: 100

Part 1: Short Answer Questions

1. Problem Definition

Hypothetical AI Problem: Predicting student dropout risk in higher education.

Objectives:

1. Identify at-risk students early.

2. Improve retention strategies.

3. Reduce dropout rates.

Stakeholders: University admin, academic advisors.

KPI: Recall score for identifying at-risk students.

2. Data Collection & Preprocessing

Data Sources: Academic records, LMS activity.

Bias: Socioeconomic factors affecting engagement.

Preprocessing: Missing data imputation, normalization, encoding.

3. Model Development

Model: Random Forest (robust, interpretable).

Split: 70/15/15.

Hyperparameters: n_estimators, max_depth.

4. Evaluation & Deployment

Metrics: Precision, Recall.

Concept drift: Changes in data patterns; monitor with drift detectors.

Deployment challenge: Scalability.

Part 2: Case Study – Hospital Readmission Prediction

Problem Scope

Predict patient readmission within 30 days.

Objectives: Reduce readmissions, improve outcomes, assist clinicians.

Stakeholders: Doctors, management, patients.

Data Strategy

Sources: EHRs, demographics, admission history.

Ethical concerns: Privacy, bias.

Pipeline: Cleaning, feature engineering, encoding, normalization.

Model Development

Model: XGBoost.

Confusion Matrix:

TP=80, FP=30, FN=20, TN=120

Precision=0.727, Recall=0.800

Deployment

Steps: Deploy API, integrate EHR, display scores, logging.

Compliance: Encryption, RBAC, audit logs, anonymization.

Optimization

Approach: Regularization to reduce overfitting.

Part 3: Critical Thinking

Ethics & Bias

Impact: Wrong risk predictions harm disadvantaged groups.

Mitigation: Stratified sampling, reweighting.

Trade-offs

Interpretability vs Accuracy: Clinicians need transparency.

Limited compute: Prefer simpler models or batch inference.

Part 4: Reflection & Workflow Diagram

Reflection

Challenge: Deployment due to compliance and integration.

Improvements: More data, hyperparameter tuning, SHAP dashboards.

Workflow Diagram:

Problem → Data Collection → Preprocessing → Model Development → Evaluation → Deployment → Monitoring.