AI Development Workflow Assignment

Part 1: Short Answer Questions (30 points)

1. Problem Definition (6 points)

- **AI Problem:** Predicting student dropout rates in secondary schools.
- **Objectives:**
    1. Identify at-risk students early.
    2. Improve retention rates through targeted interventions.
    3. Provide actionable insights to educators.
- **Stakeholders:**
    1. School administrators.
    2. Parents and guardians.
- **KPI:** Percentage reduction in dropout rates over one academic year.

2. Data Collection & Preprocessing (8 points)

- **Data Sources:**
    1. Student academic records (grades, attendance).
    2. Socioeconomic background surveys.
- **Potential Bias:** Underrepresentation of students from remote areas.
- **Preprocessing Steps:**
    1. Handle missing data via imputation.
    2. Normalize numerical features.
    3. Encode categorical variables (e.g., gender, location).

3. Model Development (8 points)

- **Model Chosen:** Random Forest

- o **Justification:** Handles both numerical and categorical data well and provides feature importance.

- **Data Split:** 70% training, 15% validation, 15% test.

- **Hyperparameters to Tune:**

1. Number of trees (n_estimators) – affects accuracy.

2. Maximum tree depth – helps prevent overfitting.

4. Evaluation & Deployment (8 points)

- **Evaluation Metrics:**

  1. F1-Score – balances precision and recall for imbalanced datasets.

  2. ROC-AUC – measures overall classification performance.

- **Concept Drift:** When model performance degrades due to changes in data patterns.

  - o **Monitoring:** Regular retraining with new data.

- **Deployment Challenge:** Ensuring scalability to handle large datasets across schools.

Part 2: Case Study Application (40 points)

Scenario: Predicting patient readmission risk within 30 days of discharge.

Problem Scope (5 points)

- **Problem:** Predict patients likely to be readmitted within 30 days.

- **Objectives:**

  1. Reduce readmission rates.

  2. Optimize post-discharge care.

- **Stakeholders:**

  1. Hospital management.

  2. Patients and caregivers.

Data Strategy (10 points)

- **Data Sources:**

    1. Electronic Health Records (EHRs).

    2. Patient demographics and medical history.

- **Ethical Concerns:**

    1. Patient privacy (data must be anonymized).

    2. Fairness (avoid bias against certain demographics).

- **Preprocessing Pipeline:**

    1. Remove PII and anonymize records.

    2. Handle missing values.

    3. Feature engineering (e.g., days since last visit, comorbidities count).

    4. Normalize lab results.

Model Development (10 points)

- **Model Selected:** Logistic Regression

    o **Justification:** Interpretable and effective for binary classification.

- **Confusion Matrix (hypothetical):**

    o TP: 60, FP: 20, FN: 15, TN: 105

- **Precision:** 60 / (60 + 20) = 0.75

- **Recall:** 60 / (60 + 15) = 0.80

Deployment (10 points)

- **Integration Steps:**

    1. Build API for model predictions.

    2. Integrate into hospital EHR system.

    3. Provide dashboard for staff alerts.

- **Compliance:** Ensure adherence to HIPAA regulations (e.g., data encryption, access control).

Optimization (5 points)

- **Method:** Use regularization (e.g., L1) to prevent overfitting.

Part 3: Critical Thinking (20 points)

Ethics & Bias (10 points)

- **Impact of Biased Data:** May result in unfair predictions, such as underestimating readmission risk for minorities.

- **Mitigation Strategy:** Use fairness-aware algorithms and diverse training data.

Trade-offs (10 points)

- **Interpretability vs. Accuracy:** Complex models (e.g., neural nets) may be more accurate but less interpretable, which is critical in healthcare.

- **Limited Resources:** Simpler models like logistic regression are preferred due to lower computational demands.

Part 4: Reflection & Workflow Diagram (10 points)

Reflection (5 points)

- **Most Challenging Part:** Preprocessing and ethical considerations due to the sensitivity of patient data.

- **Improvement:** Invest more time in validating data quality and engaging with medical experts.

Diagram (5 points)

AI Development Workflow:

[Problem Definition] --> [Data Collection] --> [Preprocessing] --> [Model Development] --> [Evaluation] --> [Deployment] --> [Monitoring & Feedback]

End of Assignment