AI Development Workflow Assignment

Part 1: Short Answer Questions

1. Problem Definition

- **Hypothetical AI Problem:** Predicting employee attrition in a large corporation.
- **Objectives:**
  1. Identify employees at high risk of leaving.
  2. Reduce overall attrition rate by 10% in one year.
  3. Provide actionable insights for HR interventions.
- **Stakeholders:**
  - Human Resources Department
  - Department Managers
- **Key Performance Indicator (KPI):**
  - Percentage reduction in actual attrition rate after model deployment.

2. Data Collection & Preprocessing

- **Data Sources:**
  1. Employee HR records (demographics, tenure, performance).
  2. Employee engagement survey results.
- **Potential Bias:**
  - Survey response bias: Employees who are dissatisfied may be less likely to complete surveys, skewing the data.
- **Preprocessing Steps:**
  1. Handle missing values (e.g., impute or remove incomplete records).
  2. Normalize numerical features (e.g., salary, tenure).
  3. Encode categorical variables (e.g., department, job role).

3. Model Development

- **Model Choice:** Random Forest – robust to overfitting, handles mixed data types, and provides feature importance.
- **Data Splitting:** 70% training, 15% validation, 15% test (stratified by attrition label).
- **Hyperparameters to Tune:**
  1. Number of trees (n_estimators): Affects model complexity and performance.
  2. Maximum tree depth (max_depth): Controls overfitting.

4. Evaluation & Deployment

- **Evaluation Metrics:**
  1. F1 Score – balances precision and recall, important for imbalanced classes.
  2. ROC-AUC – measures overall model discrimination.
- **Concept Drift:**
  - When the statistical properties of input data change over time, reducing model accuracy.
  - **Monitoring:** Track model performance metrics over time and set up alerts for significant drops.
- **Technical Challenge:**
  - Scalability – ensuring the model can handle predictions for thousands of employees in real-time.

Part 2: Case Study Application

Problem Scope

- **Problem:** Predict patient readmission risk within 30 days post-discharge.
- **Objectives:**
  - Identify high-risk patients for targeted interventions.
  - Reduce readmission rates and associated costs.
- **Stakeholders:**
  - Hospital administrators
  - Clinicians (doctors, nurses)

Data Strategy

- **Data Sources:**
  - Electronic Health Records (EHRs)
  - Patient demographics and social determinants
- **Ethical Concerns:**
  1. Patient privacy and data security.
  2. Potential for algorithmic bias affecting vulnerable groups.
- **Preprocessing Pipeline:**
  - Data cleaning (remove duplicates, handle missing values)
  - Feature engineering (e.g., count of previous admissions, comorbidity index)
  - One-hot encoding for categorical variables (e.g., discharge disposition)
  - Normalization of continuous variables (e.g., age, lab results)

Model Development

- **Model Choice:** Logistic Regression – interpretable, suitable for binary classification, and widely used in healthcare.
- **Hypothetical Confusion Matrix:**

	Predicted Readmit	Predicted No Readmit
Actual Readmit	30	10
Actual No Readmit	15	45

  - **Precision:** 30 / (30+15) = 0.67
  - **Recall:** 30 / (30+10) = 0.75

Deployment

- **Integration Steps:**
  1. Develop API for model predictions.
  2. Integrate with hospital EHR system.
  3. Train staff on model usage and interpretation.
- **Regulatory Compliance:**
  - Ensure all data handling and storage comply with HIPAA.
  - Regular audits and access controls.

Optimization

- **Overfitting Solution:** Use regularization (e.g., L1/L2 penalty in logistic regression).

Part 3: Critical Thinking

Ethics & Bias

- **Impact of Biased Data:**
  - Biased training data may lead to unfair predictions, e.g., underestimating risk for certain demographic groups, resulting in unequal care.
- **Mitigation Strategy:**
  - Implement fairness-aware algorithms and regularly audit model outputs for disparate impact.

Trade-offs

- **Interpretability vs. Accuracy:**
  - Highly accurate models (e.g., deep neural networks) may be less interpretable, which is problematic in healthcare where decisions must be explainable. Logistic regression or decision trees offer more transparency.
- **Limited Resources Impact:**
  - Resource constraints may necessitate simpler models (e.g., logistic regression) that require less computation and are easier to deploy.

Part 4: Reflection & Workflow Diagram

Reflection

- **Most Challenging Part:** Ensuring data quality and addressing bias, as healthcare data is often messy and sensitive.
- **Improvements:** With more time/resources, I would collect more diverse data and involve domain experts in feature engineering.

Workflow Diagram Description

- **Stages:**
  1. Problem Definition
  2. Data Collection
  3. Data Preprocessing
  4. Model Development
  5. Model Evaluation
  6. Deployment
  7. Monitoring & Maintenance
- *(See `diagrams/workflow_diagram.png` for the visual flowchart.)*

References

- scikit-learn documentation: https://scikit-learn.org/stable/

- pandas documentation: https://pandas.pydata.org/
- HIPAA guidelines: https://www.hhs.gov/hipaa/index.html