AI Development Workflow Assignment Report

Title: Predicting Patient Readmission within 30 Days of Discharge Using AI

Team: Amah and Co Group

Part 1: Short Answer Questions

1. Problem Definition

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

Objectives:

1. Reduce hospital readmission rates.
2. Optimize resource allocation for high-risk patients.
3. Enhance patient care quality by proactive interventions.

Stakeholders:

- Hospital administrators
- Healthcare providers (doctors, nurses)

KPI: 30-day readmission rate reduction percentage.

---

2. Data Collection & Preprocessing

Data Sources:

1. Electronic Health Records (EHRs) including diagnosis, treatment, and discharge data.

2. Patient demographics and social determinants of health (age, gender, comorbidities, socioeconomic status).

Potential Bias: Historical data may underrepresent certain patient demographics, leading to biased predictions.

Preprocessing Steps:

1. Handling missing values through imputation or removal.
2. Normalizing numeric features such as lab results.
3. Encoding categorical variables (e.g., diagnosis codes) using one-hot or label encoding.

---

3. Model Development

Model Choice: Gradient Boosting Machine (GBM) for handling structured data and providing high accuracy while managing non-linear relationships.

Data Split: 70% training, 15% validation, 15% testing to evaluate generalization.

Hyperparameters to Tune:

1. Number of trees (to balance underfitting vs overfitting).
2. Learning rate (controls model convergence speed and generalization).

---

4. Evaluation & Deployment

Evaluation Metrics:

1. Precision: Measures proportion of correctly predicted high-risk readmissions.
2. Recall: Ensures high-risk patients are correctly identified.

Concept Drift: Changes in patient population or treatment protocols may reduce model accuracy. We would monitor via periodic re-evaluation and retraining.

Technical Challenge: Integration with hospital IT systems may face scalability issues due to large EHR datasets.

Part 2: Case Study Application

Problem Scope

Problem: Predict patient readmission risk within 30 days to reduce preventable readmissions.

Objectives:

- Identify high-risk patients for proactive interventions.
- Allocate hospital resources efficiently.

Stakeholders:

- Hospital administrators
- Healthcare providers

Data Strategy

Data Sources:

- EHRs (diagnosis, lab results, treatment history)
- Patient demographics and socio-economic indicators

Ethical Concerns:

1. Patient privacy and data security.
2. Avoiding biased decisions affecting vulnerable populations.

Preprocessing Pipeline:

1. Handle missing lab values and categorical entries.
2. Normalize continuous variables (age, lab results).
3. Encode categorical features (ICD codes, treatment types).

4. Feature engineering: create risk scores from comorbidities and prior admissions.

Model Development

Model: Gradient Boosting Machine (GBM) — balances accuracy and interpretability for tabular healthcare data.

Hypothetical Confusion Matrix:

	Predicted Readmit	**Predicted No Readmit**
Actual Readmit	80	20
Actual No Readmit	15	85

Precision: 80 / (80+15) = 0.842
Recall: 80 / (80+20) = 0.8

Deployment

Integration Steps:

1. Deploy model as API endpoint linked to hospital EHR system.
2. Implement real-time predictions upon patient discharge.
3. Monitor model performance and update retraining schedule.

Compliance:

- Encrypt data at rest and in transit.
- Anonymize patient data for model development.
- Ensure adherence to HIPAA regulations.

Optimization for Overfitting:

- Implement early stopping during training.

Part 3: Critical Thinking

Ethics & Bias

Impact of Biased Data:

- Could lead to underprediction of readmission risk for certain groups, compromising patient care.

Mitigation Strategy:

- Use stratified sampling and fairness-aware algorithms to balance representation.

Trade-offs

Interpretability vs Accuracy:

- Highly complex models may increase accuracy but reduce transparency for clinicians.

Resource Constraints:

- Limited computational resources may necessitate simpler models (e.g., logistic regression) at the cost of some predictive performance.

Part 4: Reflection & Workflow Diagram

Reflection

- Most challenging part: Ensuring ethical fairness and handling biased historical data.
- Improvements: More diverse and updated datasets; cloud-based deployment for scalability.

AI Development Workflow Diagram

```
Problem Definition --> Data Collection --> Data Preprocessing -->
Model Development --> Model Evaluation --> Deployment & Monitoring
```

(Place this as a flow diagram graphic in Word/Docs for better visual appeal.)

References (APA style)

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

2. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, 380(14), 1347–1358.

3. HIPAA Journal. (2023). *HIPAA Compliance Guidelines for Healthcare AI Applications*.

4. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

5. European Commission. (2019). *Ethics Guidelines for Trustworthy AI*.