AI Assignment - Short Answer Questions

Part 1: Short Answer Questions

1. Problem Definition (6 points)

Hypothetical AI problem: "Predicting patient hospital readmission within 30 days."

3 Objectives:

1. Accurately identify patients at high risk of readmission.

2. Reduce hospital costs associated with frequent readmissions.

3. Improve patient outcomes through early interventions.

2 Stakeholders:

- Hospital administrators

- Healthcare providers (e.g., doctors, nurses)

1 Key Performance Indicator (KPI):

- Readmission rate reduction percentage

2. Data Collection & Preprocessing (8 points)

2 Data Sources:

1. Electronic Health Records (EHR) from hospital databases.

2. Insurance claim data for patient histories.

1 Potential Bias:

- Demographic bias: If certain groups (e.g., low-income or minority patients) are underrepresented or misrepresented, the model might perform poorly for them.

3 Preprocessing Steps:

1. Handle missing values in patient records.

2. Normalize continuous variables like age, length of stay, or lab results.

3. Encode categorical variables such as diagnosis codes or treatment types.

3. Model Development (8 points)

Chosen Model & Justification:

- Random Forest: It works well for tabular healthcare data, handles non-linear relationships, is interpretable (feature importance), and reduces overfitting.

Data Splitting:

- 70% for training, 15% for validation, 15% for testing.

2 Hyperparameters to Tune:

1. Number of trees: Balances model performance and computational cost.

2. Maximum tree depth: Prevents overfitting by limiting how deep each tree can grow.

4. Evaluation & Deployment (8 points)

2 Evaluation Metrics & Relevance:

1. AUC-ROC score: Good for measuring how well the model distinguishes between readmission vs. non-readmission cases.

2. F1-score: Balances precision and recall - important if there is an imbalance.

Concept Drift:

- Concept drift occurs when the statistical properties of input data change over time.

- Monitoring: Regularly compare recent predictions with actual outcomes; retrain the model if drift is detected.

1 Technical Challenge during Deployment:

- Scalability: Ensuring the model can handle real-time predictions for thousands of patients.