AI Assignment - Part 1: Short Answer Questions

1. Problem Definition (6 points)

Hypothetical AI Problem:
Predicting Student Failure Rates in a Specific Course

This AI problem focuses on identifying students who are at risk of failing a particular course using historical and behavioral data.

Objectives:
1. To determine the average success and failure rates in the course.
2. To identify early signs of potential failure and suggest remedial actions.
3. To recommend improved teaching or learning strategies for better academic outcomes.

Stakeholders:
- Primary Stakeholders: Students and the Lecturer teaching the course
- Secondary Stakeholders: Department heads, university administration, and parents/guardians of the students

Key Performance Indicator (KPI):
- Success Rate of students in the course (i.e., percentage of students who pass after intervention)

2. Data Collection & Preprocessing (8 points)

Data Sources:
1. Lecturer's Datasheet - containing scores, attendance records, assignment marks, and class participation
2. Student Surveys - gathering perceptions on course difficulty, teaching methods, and self-assessed understanding

Potential Bias in the Data:
- Response Bias: Student feedback may reflect subjective opinions or personal biases about the lecturer rather than objective performance indicators.

Recommended Data Fields:
- Attendance records
- Assignment and test participation
- Lecturer consistency and availability
- Classroom size
- Attentiveness and punctuality
- Tutorial attendance
- Previous academic records

AI Assignment - Part 1: Short Answer Questions

Preprocessing Steps:
1. Handling Missing Data - filling in missing values such as incomplete test scores or skipped survey items
2. Normalization - scaling features like attendance percentage and scores to ensure consistent value ranges
3. Encoding Categorical Variables - converting categorical features (e.g., gender, department) into numerical format for the model

3. Model Development (8 points)

Selected Model and Justification:
- Logistic Regression
  Chosen for its simplicity, interpretability, and effectiveness in binary classification tasks like predicting pass/fail outcomes.

Alternatives like Random Forest or XGBoost can also be explored if better performance is needed.

Data Splitting Strategy:
- Training Set: 70%
- Validation Set: 15%
- Test Set: 15%

The validation set helps with model tuning, while the test set evaluates final performance.

Two Hyperparameters to Tune:
1. Regularization Strength (C in Logistic Regression) - to prevent overfitting by controlling the complexity of the model
2. Learning Rate (in gradient-based models) - determines how quickly the model adjusts to errors during training

4. Evaluation & Deployment (8 points)

Evaluation Metrics:
1. Accuracy - the proportion of total correct predictions; useful when classes are balanced
2. F1 Score - the harmonic mean of precision and recall; better when failure cases are less frequent and more critical to detect

Concept Drift and Monitoring:
- Concept Drift refers to changes in the underlying data patterns over time that affect model accuracy (e.g., new teaching methods or curriculum changes).
- To monitor drift:
  - Track performance metrics like accuracy over time
  - Periodically compare distributions of recent and past data

AI Assignment - Part 1: Short Answer Questions

- Retrain the model with fresh data if performance degrades

Technical Challenge During Deployment:

- Model Integration: Ensuring the predictive model can be embedded seamlessly into existing school or university systems (e.g., learning management systems).

Other potential challenges include:

- Scalability - Can the model support large volumes of data?
- Data Privacy Compliance - Ensuring student data is protected under relevant policies.