___

1. Problem Definition (6 points)

Hypothetical AI Problem:

Predicting student dropout rates in a university.

Objectives (3):

1. Identify students at high risk of dropping out early.
2. Provide timely intervention strategies for at-risk students.
3. Improve overall student retention and academic performance.

Stakeholders (2):

1. University administration.
2. Students.

Key Performance Indicator (1):

- KPI: Retention rate improvement (percentage decrease in dropout rates after model implementation).

2. Data Collection & Preprocessing (8 points)

Data Sources (2):

1. Student academic records (grades, attendance).
2. Student support service logs (counseling and

engagement reports).

Potential Bias (1):

- Students from economically disadvantaged backgrounds may be overrepresented in the dropout category, causing the model to unfairly label similar students as high-risk.

Preprocessing Steps (3):

1. Handle missing values (e.g., fill with median or remove incomplete entries).
2. Normalize/standardize numerical features (e.g., attendance percentages).
3. Encode categorical variables (e.g., program of study, housing status).

3. Model Development (8 points)

Chosen Model: Random Forest
Justification: It handles mixed data types well, is robust to noise, and gives feature importance insights.

Data Splitting Strategy:

- 70% training, 15% validation, 15% test set.

Hyperparameters to Tune (2):

1. Number of trees (n_estimators): Controls model

complexity and performance.

2. Maximum depth: Prevents overfitting by limiting tree growth.

———

4. Evaluation & Deployment (8 points)

Evaluation Metrics (2):

1. Accuracy: Measures overall correct predictions.
2. F1-Score: Useful when classes are imbalanced (e.g., fewer dropouts).

Concept Drift:

- Concept drift occurs when the patterns in the data change over time, causing the model ' s performance to degrade.
- Monitoring: Continuously track evaluation metrics and periodically retrain the model using recent data.

Deployment Challenge (1):

- Scalability: Serving predictions to thousands of students at peak registration times may require optimized infrastructure or cloud scaling.