Part 1:

3. Model Development (Group Member 3)

Chosen Model: Random Forest

Justification:

Handles both numerical and categorical data well.

Offers feature importance for interpretability.

Data Splits:

Training: 70%

Validation: 15%

Test: 15%

Hyperparameters:

Number of trees (n\_estimators): Impacts model accuracy.

Maximum depth (max\_depth): Controls overfitting.

4. Evaluation & Deployment (Group Member 3)

Evaluation Metrics:

Precision: Critical when false positives (unnecessary intervention) are costly.

Recall: Important to minimize false negatives (missed dropouts).

Concept Drift:

Model performance degrades as student behaviors change over time.

Monitoring Strategy:

Periodic re-evaluation and retraining using recent data.

Deployment Challenge:

Scalability—model must process large volumes of student data across platforms.

Part 2:

Model Development (Group Member 3)

Model Chosen: Logistic Regression

Justification:

Interpretable and commonly used in healthcare for risk prediction.

Confusion Matrix (Hypothetical):

Predicted: No Readmission Predicted: Readmission

Actual: No 720 80

Actual: Yes 50 150

Precision: 150 / (150 + 80) = 0.652

Recall: 150 / (150 + 50) = 0.75