BRAINWARE UNIVERSITY

Term End Examination 2025-2026
Programme – M.Tech.(CSE)-AIML-2025
Course Name – Explainable AI (XAI)
Course Code - MTA20406
(Semester II)

Full Marks : 60 **Time : 2:30 Hours**

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group-A

(Multiple Choice Type Question) 1 x 15=15

1. *Choose the correct alternative from the following :*

(i) Recognize an advantage of intrinsic explainability.
   a) External justification b) Transparency by design
   c) Post-processing d) Increased opacity

(ii) Recognize a benefit of explainability in healthcare.
   a) Error detection b) Automation
   c) Data expansion d) Visualization

(iii) Analyze a limitation of black-box models in regulated domains.
   a) Speed b) Lack of accountability
   c) Data size d) Automation

(iv) Apply SHAP values to interpret feature contribution.
   a) Random weighting b) Game-theoretic attribution
   c) Gradient masking d) Rule extraction

(v) Use SHAP to obtain global feature importance.
   a) Single-instance SHAP b) Aggregated SHAP values
   c) Random sampling d) Rule induction

(vi) Analyze an assumption underlying Partial Dependence Plots.
   a) Feature independence b) Feature balance
   c) Label symmetry d) Data normalization

(vii) Use feature importance to interpret a tree-based model.
   a) Measure split contribution b) Compute gradients
   c) Generate anchors d) Train surrogate CNN

(viii) Implement attention visualization for neural network interpretability.
   a) Loss curves b) Attention weight maps
   c) Feature shuffling d) PDP plots

(ix) Analyze the role of anchors in explanation reliability.

a) Low precision rules
b) High-precision local explanations
c) Global feature ranking
d) Dataset summarization

(x) Explain accountability requirements in ethical AI.

a) Performance tuning
b) Traceability and justification
c) Feature pruning
d) Data labeling

(xi) Evaluate disparate impact as a fairness metric.

a) Comprehensive
b) Limited to outcomes
c) Perfect accuracy
d) Model-specific

(xii) Formulate a compliance-oriented explainability approach.

a) Model hiding
b) Transparency mechanisms
c) Data encryption
d) Compression

(xiii) Apply ELI5 to interpret a healthcare diagnostic model output.

a) Feature attribution
b) Data normalization
c) Model retraining
d) Dataset balancing

(xiv) Apply XAI methods to interpret medical image classification models.

a) Attribution visualization
b) Noise injection
c) Model compression
d) Image resizing

(xv) Use Captum attribution to support regulatory transparency.

a) Feature influence analysis
b) Encryption
c) Compression
d) Obfuscation

Group-B

(Short Answer Type Questions) 3 x 5=15

2. Implement a surrogate model to explain a complex tree-based ensemble. (3)
3. Critique the limitations of disparate impact as a sole fairness metric. (3)
4. List two regulatory frameworks that emphasize explainability in AI. (3)
5. Calculate permutation importance when accuracy drops from 90% to 85% after feature shuffling. (3)
6. Design an explainable reinforcement learning framework for autonomous decision-making. (3)

OR

Construct an XAI workflow using InterpretML for transparent credit scoring. (3)

Group-C

(Long Answer Type Questions) 5 x 6=30

7. Describe the concept of algorithmic accountability in regulated AI environments. (5)
8. Analyze the interpretability advantages of rule-based systems using decision rules. (5)
9. Analyze the importance of accountability in explainable AI-driven decisions. (5)
10. Apply SHAP to compute and interpret feature contributions for a model prediction. (5)
11. Analyze the strengths and limitations of LIME as a model-agnostic explanation technique. (5)
12. Design an explainable AI framework for interpreting healthcare diagnostic models. (5)

OR

Construct a transparent credit scoring pipeline using XAI tools. (5)

****