Quiz: Detection Pipeline Framework

Section 03 – Digital-AI-Finance

Joerg Osterrieder

Zurich University of Applied Sciences (ZHAW)

2025

Question 1: Pipeline Stages

How many stages does the detection pipeline have?

a) 3
b) 4
c) 5
d) 6

Question 1: Pipeline Stages

How many stages does the detection pipeline have?

a) 3
b) 4
c) 5
d) 6

Answer

c) 5
The detection pipeline comprises five sequential stages: data ingestion and integration, feature engineering, model selection and training, explainability and interpretation, and deployment and monitoring.

Source: Section 3.1

Question 2: Feature Families

How many feature families are identified?

a) 3
b) 4
c) 5
d) 7

Question 2: Feature Families

How many feature families are identified?

a) 3
b) 4
c) 5
d) 7

Answer

c) 5
Five feature families are identified: statistical features, Benford's law features, textual features from regulatory filings, network and relational features, and temporal features.

Source: Section 3.3

Question 3: Gradient Boosting Performance

What F_1 score do gradient boosting ensembles approach?

a) 0.72
b) 0.78
c) 0.84
d) 0.88

Question 3: Gradient Boosting Performance

What F_1 score do gradient boosting ensembles approach?

a) 0.72
b) 0.78
c) 0.84
d) 0.88

Answer

d) 0.88
Recent work on stacking ensembles that combine XGBoost, LightGBM, and CatBoost through a meta-learner has reported F_1 scores approaching 0.88, the highest among individual method families.

Source: Section 3.4

Question 4: Deep Autoencoder AUC

What AUC do deep autoencoders achieve for anomaly detection?

a) 0.65
b) 0.72
c) 0.79
d) 0.85

Question 4: Deep Autoencoder AUC

What AUC do deep autoencoders achieve for anomaly detection?

a) 0.65
b) 0.72
c) 0.79
d) 0.85

Answer

c) 0.79
Deep autoencoders have achieved an AUC of approximately 0.79 on hedge fund return data in anomaly detection frameworks, competitive with supervised approaches despite requiring no fraud labels.

Source: Section 3.4

Question 5: SHAP Meaning

What does SHAP stand for?

a) Statistical Hypothesis Analysis Protocol
b) SHapley Additive exPlanations
c) Systematic Hedge fund Assessment Program
d) Supervised Heuristic Analysis Pipeline

Question 5: SHAP Meaning

What does SHAP stand for?

a) Statistical Hypothesis Analysis Protocol
b) SHapley Additive exPlanations
c) Systematic Hedge fund Assessment Program
d) Supervised Heuristic Analysis Pipeline

Answer

b) SHapley Additive exPlanations
SHAP (SHapley Additive exPlanations) values provide theoretically grounded feature attribution scores that decompose a model's prediction into the contribution of each input feature.

Source: Section 3.5

Question 6: Fusion Approaches

Which fusion approach is better when data sources have different scales?

a) Early fusion
b) Late fusion
c) Both equal
d) Neither works

Question 6: Fusion Approaches

Which fusion approach is better when data sources have different scales?

a) Early fusion
b) Late fusion
c) Both equal
d) Neither works

Answer

b) Late fusion
Late fusion processes each data source through a modality-specific feature extraction pipeline and combines results afterward, preserving native structure and accommodating different scales better than early fusion.

Source: Section 3.2

Question 7: Feedback Loops

What are the two feedback loops in the pipeline?

a) Training and testing
b) Investigator feedback and drift-triggered retraining
c) Input and output validation
d) Forward and backward propagation

Question 7: Feedback Loops

What are the two feedback loops in the pipeline?

a) Training and testing
b) Investigator feedback and drift-triggered retraining
c) Input and output validation
d) Forward and backward propagation

Answer

b) Investigator feedback and drift-triggered retraining
The pipeline includes two feedback loops from deployment back to earlier stages: investigator feedback to feature engineering and drift-triggered retraining to model selection.

Source: Section 3.1