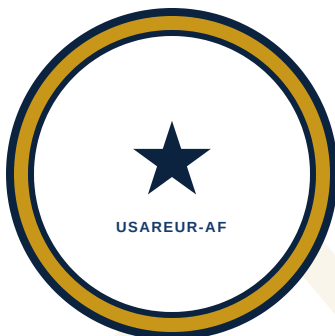


DRAFT — UNOFFICIAL — NOT FOR OPERATIONAL USE

PRACTICAL EXERCISE

EX-40M



EX_40M — ML Engineer

Practical Exercise — SL 4M Proficiency

HEADQUARTERS
UNITED STATES ARMY EUROPE AND AFRICA
(USAREUR-AF)
Wiesbaden, Germany

DRAFT — NOT FOR OFFICIAL USE. FOR TRAINING PLANNING PURPOSES ONLY.

26 MARCH 2026

DRAFT — UNOFFICIAL — NOT FOR OPERATIONAL USE

EX_40M — ML ENGINEER

PRACTICAL EXERCISE — SL 4M PROFICIENCY

Field	Value
Version	1.0 — March 2026
Prerequisite	SL 3 REQUIRED; SL 4M — Machine Learning Engineer Technical Manual (and SL 1 through SL 2)
Duration	3–4 hours
Environment	MSS with Python Transforms, Model Integration enabled — see ENVIRONMENT_SETUP.md

COMPANION RESOURCES

Resource	Reference
Technical Manual	SL 4M — ML Engineer
Syllabus	SYLLABUS_TM40M
Pre-Exercise Exam	EXAM_TM40M_PRE
Post-Exercise Exam	EXAM_TM40M_POST
Continuation Track	SL 5M — Advanced ML Engineer

WFF AWARENESS

The NMC risk model produced in this exercise is an operational product consumed by WFF track personnel (SL 4D Sustainment in particular) as a readiness decision-support tool. Evaluators should verify that the model card (Task 5) is written for a non-MLE audience — sustainment officers and NCOs will be the primary consumers.

SCENARIO

The OPDATA team requires a predictive model that flags equipment at risk of NMC (non-mission-capable) status within the next 7 days based on historical maintenance records. Build a classification model, validate it, deploy it to the Ontology, and document it for governance.

Training dataset: Synthetic PMCS records with NMC labels, 12 months, ~5,000 equipment events.

TASK LIST

Task 1 — Feature Engineering (45 min)

- Build a feature pipeline (Python Transform) with the following features: days since last service, maintenance frequency (30/60/90 day windows), equipment class one-hot encoding, unit
- Handle nulls with a documented imputation strategy
- Output feature matrix with target label `nmc_within_7d`

Standard	Criteria
Go	Feature matrix builds; no nulls in output; all features present
No-Go	Transform errors or nulls remain without documentation

Task 2 — Model Training and Selection (60 min)

- Train at least 2 classifiers (e.g., logistic regression + gradient boosted tree)
- Evaluate on a held-out test set: report precision, recall, F1, and AUC-ROC
- Select the better model and justify the choice in writing (2–3 sentences)

Standard	Criteria
Go	Both models trained; metrics reported from test set; selection justified
No-Go	Only one model trained or metrics not reported

Task 3 — Validation and Failure Mode Analysis (30 min)

- Produce a confusion matrix for the selected model
- Identify the dominant failure mode (false positives or false negatives) and explain its operational consequence

- Define a decision threshold appropriate for the operational context — justify whether to err toward false positive or false negative

Standard	Criteria
Go	Confusion matrix correct; failure mode identified; threshold justified with operational reasoning
No-Go	Confusion matrix wrong or threshold unjustified

Task 4 — Deploy to Ontology (30 min)

- Register the model in Foundry Model Integration
- Create an Object Type property: `nmc_risk_score` backed by the model
- Verify the property populates for a sample equipment object

Standard	Criteria
Go	Property appears on equipment objects with model-backed values
No-Go	Property does not populate or model registration fails

Task 5 — Model Card (20 min)

Write a model card covering all six required sections: purpose, training data, features, performance metrics, known limitations, refresh cadence.

Standard	Criteria
Go	All six sections present; limitations section is honest
No-Go	Limitations section absent or metrics contradict Task 2 results

EVALUATOR NOTES

Scoring: 5 tasks. Go on 4 of 5 = overall Go. No-Go on Task 3 or Task 4 = automatic No-Go (validation rigor and production deployment are the core MLE competencies).

Pre-Exercise Checklist

- Confirm Python Transforms are enabled for training accounts
- Confirm Model Integration (Foundry model registry) is enabled
- Confirm the `Equipment` Object Type exists in the training Ontology with a `nmc_risk_score` property slot

- Confirm training accounts can write model-backed properties (requires specific permission in some tenants)

Common Failure Modes

Task	Common Failure	Evaluator Guidance
Task 1	Nulls remain in output	Ask participant to show their imputation step; undocumented imputation = No-Go
Task 2	Only one model trained	Automatic No-Go; remind participant the task requires at least 2 classifiers
Task 2	Metrics reported from training set not test set	Ask: "What was your test set size?" — if no holdout, No-Go
Task 3	Threshold not justified with operational reasoning	Generic "use 0.5" without justification = No-Go; must reference false positive/negative consequence
Task 4	Model registered but property does not populate	Usually a permission or property type issue — allow 15 min troubleshooting before marking No-Go
Task 5	Limitations section absent or optimistic	The dataset has known class imbalance (~10% positive rate) — if not mentioned, ask directly

Timing Notes

- Task 2 (model training) averages 60–90 min depending on model complexity; gradient boosted trees will take longer
- Task 3 (confusion matrix and threshold) is conceptually straightforward but participants often spend significant time on the operational consequence reasoning — this is intentional
- Consider splitting EX_40M across two sessions if time is constrained

NEXT STEPS

Participants who receive an overall Go on EX_40M are eligible to enroll in **SL 5M — Advanced ML Engineer**. SL 5M extends into model retraining pipelines, drift detection, and MLOps governance for production Foundry environments. No TM-50A–F tracks exist.

ENVIRONMENT SETUP

See [ENVIRONMENT_SETUP.md](#) for full setup instructions.