

AWS ML Associate — 1-Page Exam Cheat Sheet

Section 1: Data Engineering

- 1.1 Collect/Ingest/Store → S3 hub; Glue ETL; Kinesis (streams); DMS (DBs); Transfer Family (SFTP); Snowball (bulk).
- 1.2 Prep Data → Glue DataBrew (no-code), Data Wrangler (preferred), Processing jobs, Feature Store.
- 1.3 Train Models → Built-in algos (XGBoost, Linear Learner), Pre-built containers (TF/PyTorch), Autopilot baselines.

Section 2: Modeling

- 2.1 Choose Approach → Classification (categorical), Regression (continuous), Clustering (K-Means), Recs (Factorization Machines).
- 2.2 Feature Eng. → One-hot, embeddings, scaling (min-max, z-score), PCA for dimension reduction.
- 2.3 Train/Eval → Split data; Metrics: Class (Acc, Prec/Rec, F1, ROC-AUC), Regr (RMSE, MAE), Cluster (Silhouette).
- 2.4 Tune → Hyperparams; Random/Grid/Bayesian; Auto Model Tuning in SageMaker.

Section 3: Deployment & Ops

- 3.1 Deploy → Real-time (low latency), Batch (offline), Async (long jobs), Neo (edge).
- 3.2 Monitor → Model Monitor (drift), Clarify (bias/explainability), CloudWatch (metrics/logs).
- 3.3 Automate → Pipelines (ML-native CI/CD), Step Functions (orchestration), CodePipeline (integrations).

Section 4: Responsible, Secure, Optimize

- 4.1 Responsible AI → Clarify (bias, SHAP explainability).
- 4.2 Secure → Encrypt (SSE-S3/KMS), IAM least privilege, VPC isolation.
- 4.3 Optimize → Spot instances (checkpoint), Debugger (profiling), right-size instances, distributed training.

Mnemonic: S3 hub → Glue → Wrangler → SageMaker → Deploy → Monitor → Automate → Secure & Optimize.