

Fiche de Présentation : AWS Certified Machine Learning

Cette fiche détaille les services et notions AWS cruciaux à maîtriser pour la certification **AWS Certified Machine Learning - Specialty (MLS-C01)**. Cette certification valide votre expertise dans la conception, l'implémentation, le déploiement et la maintenance de solutions de Machine Learning (ML) sur la plateforme AWS.

I. Services AWS Fondamentaux pour le Machine Learning

1. Amazon SageMaker

Plateforme centrale et entièrement gérée pour le cycle de vie du ML.

- **Instances de notebook SageMaker** : Environnements Jupyter/JupyterLab pour l'exploration de données et le développement de modèles.
- **SageMaker Studio** : IDE web unifié pour toutes les étapes du ML (notebooks, débogage, suivi d'expériences, etc.).
- **SageMaker Ground Truth** : Service d'étiquetage de données pour créer des jeux de données d'entraînement de haute qualité (humain ou automatique).
- **SageMaker Data Wrangler** : Outil de préparation et de visualisation de données pour l'ingénierie des caractéristiques.
- **SageMaker Feature Store** : Répertoire centralisé pour stocker, découvrir, partager et gérer les caractéristiques (features) pour l'entraînement et l'inférence.
- **Traitement SageMaker (Processing Jobs)** : Exécution de tâches de prétraitement, post-traitement et évaluation de modèles dans un environnement géré.
- **Entraînement SageMaker (Training Jobs)** :
 - **Algorithmes intégrés** : XGBoost, Linear Learner, K-Means, PCA, DeepAR, BlazingText, etc.
 - **Frameworks supportés** : TensorFlow, PyTorch, MXNet, Scikit-learn, etc. (scripts personnalisés).
 - **Bring Your Own Algorithm/Container (BYOA/BYOC)** : Utilisation de vos propres algorithmes ou conteneurs Docker.
 - **Réglage automatique des hyperparamètres (Hyperparameter Tuning / HPO)** : Optimisation automatisée des hyperparamètres du modèle.
 - **Entraînement distribué** : Pour les grands jeux de données et modèles complexes.
- **Inférence SageMaker (Inference)** :
 - **Points de terminaison en temps réel (Real-time Endpoints)** : Pour des prédictions à faible latence et haut débit.

- **Inférence par lots (Batch Transform)** : Pour des prédictions sur de grands volumes de données stockées.
- **SageMaker Serverless Inference** : Option d'inférence sans serveur, pay-as-you-go, pour les charges de travail intermittentes.
- **Multi-Model Endpoints (MME)** : Hébergement de plusieurs modèles sur un seul point de terminaison.
- **SageMaker Edge Manager** : Optimisation et gestion de modèles sur des appareils en périphérie.
- **SageMaker Pipelines** : Outil de CI/CD pour orchestrer et automatiser les workflows de ML.
- **SageMaker Model Registry** : Catalogue pour versionner, gérer et approuver les modèles avant le déploiement.
- **SageMaker Model Monitor** : Surveillance des modèles en production pour détecter la dérive des données (data drift) et la dérive de la qualité du modèle (model quality drift).
- **SageMaker Clarify** : Détection de biais dans les données et les modèles, et explicabilité des prédictions (SHAP).
- **SageMaker JumpStart** : Accès rapide à des solutions ML pré-construites, des modèles pré-entraînés et des notebooks d'exemple.

2. Services de Stockage et de Gestion de Données

- **Amazon S3 (Simple Storage Service)** :
 - Stockage d'objets principal pour les jeux de données brutes et traitées, les scripts, les artefacts de modèles.
 - Classes de stockage (Standard, Intelligent-Tiering, Glacier).
 - Gestion des versions, politiques de cycle de vie.
 - Sécurité (ACLs, politiques de bucket, chiffrement).
- **AWS Glue** : Service ETL (Extract, Transform, Load) entièrement géré.
 - **Glue Data Catalog** : Référentiel de métadonnées centralisé.
 - **Glue Crawlers** : Découverte automatique de schémas.
 - **Glue ETL Jobs (Python Shell, Apache Spark)** : Transformation et préparation de données.
 - **Glue DataBrew** : Outil visuel de préparation de données sans code.
- **Amazon Athena** : Requêtage interactif de données dans S3 en utilisant SQL standard.
- **Amazon Redshift** : Entrepôt de données (Data Warehouse) pour l'analyse de données structurées à grande échelle.
- **Amazon Kinesis** : Services pour la collecte, le traitement et l'analyse de données en streaming.

- **Kinesis Data Streams** : Ingestion de flux de données en temps réel.
- **Kinesis Data Firehose** : Chargement de flux de données vers S3, Redshift, Elasticsearch, etc.
- **Kinesis Data Analytics** : Analyse de flux de données avec SQL ou Apache Flink.
- **AWS Lake Formation** : Création, sécurisation et gestion simplifiées de lacs de données (data lakes).
- **Amazon DynamoDB** : Base de données NoSQL clé-valeur et de documents pour une faible latence (ex: stockage de caractéristiques en ligne, métadonnées).

3. Services de Calcul et de Conteneurisation

- **Amazon EC2 (Elastic Compute Cloud)** :
 - Capacité de calcul redimensionnable (instances CPU, GPU, optimisées pour la mémoire).
 - Peut être utilisé pour des tâches ML personnalisées ou l'auto-hébergement de modèles (moins courant avec SageMaker).
- **AWS Lambda** : Calcul sans serveur pour exécuter du code en réponse à des événements (ex: déclencher des pipelines, prétraitement léger, invocation de points de terminaison).
- **Amazon ECR (Elastic Container Registry)** : Registre de conteneurs Docker pour stocker et gérer les images utilisées par SageMaker (BYOC).
- **Amazon ECS (Elastic Container Service) & EKS (Elastic Kubernetes Service)** : Services d'orchestration de conteneurs. Peuvent être utilisés pour des déploiements ML, mais SageMaker est souvent préféré pour sa spécialisation.

4. Services d'IA de Haut Niveau (AI Services)

Ces services fournissent des capacités d'IA pré-entraînées via des API, utiles pour intégrer rapidement l'IA dans les applications sans gérer l'infrastructure de ML sous-jacente.

- **Vision** :
 - **Amazon Rekognition** : Analyse d'images et de vidéos (détection d'objets, reconnaissance faciale, analyse de texte dans les images).
- **Langage** :
 - **Amazon Comprehend** : Traitement du langage naturel (NLP) (analyse de sentiments, extraction d'entités, détection de langue).
 - **Amazon Translate** : Traduction automatique.
 - **Amazon Polly** : Synthèse vocale (texte en parole).
 - **Amazon Transcribe** : Reconnaissance vocale (parole en texte).
 - **Amazon Lex** : Création d'interfaces conversationnelles (chatbots).

- **Amazon Kendra** : Service de recherche d'entreprise intelligent basé sur le ML.
- **Amazon Textract** : Extraction de texte et de données à partir de documents numérisés.
- **Recommandation et Prévion** :
 - **Amazon Personalize** : Création de systèmes de recommandation personnalisés.
 - **Amazon Forecast** : Prévion de séries temporelles basée sur le ML.
- **Autres** :
 - **Amazon Fraud Detector** : Détection de fraudes en ligne.
 - **Amazon CodeGuru** : (Moins directement ML) Recommandations pour améliorer la qualité du code et les performances des applications.

5. Services de Sécurité, Surveillance et Orchestration

- **AWS IAM (Identity and Access Management)** :
 - Gestion des utilisateurs, groupes, rôles et politiques pour contrôler l'accès aux ressources AWS.
 - Crucial pour sécuriser les données, les modèles et les opérations SageMaker (principe du moindre privilège).
- **AWS KMS (Key Management Service)** : Gestion des clés de chiffrement pour protéger les données au repos et en transit.
- **Amazon CloudWatch** :
 - Surveillance des métriques et des logs des services AWS (ex: instances SageMaker, points de terminaison, tâches d'entraînement).
 - Configuration d'alarmes pour les performances et les erreurs.
- **AWS CloudTrail** : Enregistrement des appels d'API AWS pour l'audit, la sécurité et le dépannage.
- **AWS Step Functions** : Orchestration de workflows sans serveur, y compris des pipelines ML complexes impliquant plusieurs services AWS.
- **AWS Secrets Manager** : Gestion sécurisée des secrets (clés d'API, identifiants de base de données).
- **Amazon VPC (Virtual Private Cloud)** : Isolation de vos ressources AWS dans un réseau virtuel. Configuration de points de terminaison VPC pour SageMaker pour un accès sécurisé.

II. Notions Clés en Machine Learning et Spécifiques à AWS

1. Concepts Fondamentaux du Machine Learning

- **Types d'apprentissage** : Supervisé (classification, régression), non supervisé (clustering, réduction de dimension), par renforcement.

- **Cycle de vie du ML** : Définition du problème, collecte et préparation des données, ingénierie des caractéristiques, sélection et entraînement du modèle, évaluation, déploiement, surveillance et maintenance.
- **Ingénierie des caractéristiques (Feature Engineering)** :
 - Nettoyage des données (valeurs manquantes, aberrantes).
 - Transformation (normalisation, standardisation, mise à l'échelle).
 - Encodage (one-hot, label encoding).
 - Création de caractéristiques (polynomiales, interactions).
 - Sélection de caractéristiques (filtrage, wrapper, méthodes intégrées).
- **Algorithmes courants et leurs cas d'usage** :
 - Linéaires : Régression linéaire, Régression logistique.
 - Basés sur les arbres : Arbres de décision, Random Forest, Gradient Boosting (XGBoost, LightGBM, CatBoost).
 - Autres : SVM, K-Nearest Neighbors (KNN), Naive Bayes.
 - Clustering : K-Means, DBSCAN, Hiérarchique.
 - Réduction de dimension : PCA, t-SNE.
 - Deep Learning : Concepts de base des réseaux de neurones (ANN, CNN, RNN, Transformers), fonctions d'activation, optimisation (SGD, Adam), backpropagation.
- **Évaluation des modèles** :
 - **Classification** : Matrice de confusion, exactitude (accuracy), précision, rappel (recall), F1-score, AUC-ROC, AUC-PR, log loss.
 - **Régression** : MSE, RMSE, MAE, R^2 .
 - Techniques de validation : Validation croisée (k-fold), hold-out.
- **Surapprentissage (Overfitting) et Sous-apprentissage (Underfitting)** :
 - Identification et techniques de mitigation (régularisation L1/L2, dropout, augmentation de données, early stopping, simplification du modèle).
- **Réglage des hyperparamètres** : Grid search, random search, optimisation bayésienne.
- **Interprétabilité et Explicabilité** : Importance et méthodes (SHAP, LIME).
- **Détection de biais et équité (Fairness)** : Comprendre les sources de biais et comment les mesurer/mitiger.
- **Dérive des données (Data Drift) et Dérive des concepts (Concept Drift)** : Leur impact et comment les surveiller.

2. Aspects Opérationnels et Spécifiques à AWS (MLOps)

- **Sécurité des solutions ML sur AWS** :
 - Chiffrement des données au repos (S3, EBS avec KMS) et en transit (TLS/SSL).

- Utilisation des rôles IAM avec le principe du moindre privilège.
- Sécurisation des points de terminaison SageMaker (authentification IAM, VPC endpoints).
- Isolation réseau (VPC, sous-réseaux, groupes de sécurité).
- **Optimisation des coûts :**
 - Choix des types d'instances appropriés (CPU vs GPU, instances Spot pour l'entraînement).
 - Gestion du cycle de vie des données S3.
 - Arrêt des ressources inutilisées (notebooks, points de terminaison).
 - Utilisation de SageMaker Savings Plans.
 - Optimisation des inférences (ex: SageMaker Inference Recommender, Serverless Inference).
- **Haute disponibilité, scalabilité et tolérance aux pannes :**
 - Déploiement multi-AZ pour les points de terminaison SageMaker.
 - Auto Scaling pour les points de terminaison.
 - Sauvegarde et versionnement des modèles et des données.
- **Automatisation et MLOps :**
 - CI/CD pour les pipelines ML (SageMaker Pipelines, AWS CodePipeline, CodeCommit, CodeBuild).
 - Reproductibilité des expériences.
 - Gestion des versions de code, de données et de modèles.
- **Formats de données pour SageMaker :** CSV, RecordIO-protobuf, LibSVM, JSONLines. Comprendre comment les algorithmes SageMaker attendent les données.
- **Utilisation de conteneurs Docker avec SageMaker :** Structure des conteneurs pour l'entraînement et l'inférence (BYOC).
- **Surveillance et Journalisation :** Utilisation de CloudWatch Logs et Metrics pour suivre les performances et diagnostiquer les problèmes.