

Guia do exame AWS Certified Machine Learning - Specialty (MLS-C01)

Introdução

O exame AWS Certified Machine Learning - Specialty (MLS-C01) é destinado a indivíduos que desempenham uma função de desenvolvimento de inteligência artificial/machine learning (IA/ML) ou ciência de dados. O exame valida a capacidade de um candidato de projetar, criar, implantar, otimizar, treinar, ajustar e manter soluções de ML para determinados problemas de negócios usando a nuvem AWS.

O exame também valida a capacidade de um candidato de concluir as seguintes tarefas:

- Selecionar e justificar a abordagem de ML adequada para determinado problema de negócios
- Identificar os serviços da AWS apropriados para implementar soluções de ML
- Projetar e implementar soluções de ML dimensionáveis, econômicas, confiáveis e seguras

OBSERVAÇÃO: o exame AWS Certified Machine Learning - Specialty está disponível atualmente em inglês, japonês, chinês simplificado e coreano.

Descrição do candidato

Espera-se que o candidato tenha no mínimo dois anos de experiência prática em desenvolvimento, arquitetura e execução de cargas de trabalho de ML ou deep learning na nuvem AWS.

Conhecimento da AWS recomendado

O candidato deve demonstrar conhecimento sobre:

- Capacidade de expressar a intuição por trás de algoritmos básicos de ML
- Experiência na execução de otimização básica de hiperparâmetros
- Experiência com frameworks de ML e deep learning
- Capacidade de seguir as práticas recomendadas de treinamento de modelos
- Capacidade de seguir as práticas recomendadas de implantação
- Capacidade de seguir as práticas operacionais recomendadas

O que é considerado fora do escopo do candidato?

Veja a seguir uma lista (não completa) de tarefas de trabalho relacionadas as quais não se espera que o candidato seja capaz de executar. Estes itens são considerados fora do escopo do exame:

- Desenvolvimento de algoritmos extensos ou complexos
- Extensa otimização de hiperparâmetros
- Provas e cálculos matemáticos complexos
- Redes avançadas e design de rede
- Conceitos avançados de banco de dados, segurança e DevOps
- Tarefas relacionadas a DevOps para o Amazon EMR

Para obter uma lista detalhada de ferramentas e tecnologias específicas que podem ser abordadas no exame, bem como listas de serviços da AWS que estão dentro e fora do escopo, consulte o Apêndice.

Conteúdo do exame

Tipos de respostas

Existem dois tipos de perguntas no exame:

- **Múltipla escolha:** há uma resposta correta e três respostas incorretas (distratores)
- **Múltipla resposta:** há duas ou mais respostas corretas dentre cinco ou mais opções de respostas

Selecione uma ou mais respostas que completem melhor a declaração ou respondam à pergunta. Distratores, ou respostas incorretas, são opções de resposta que um candidato com habilidades ou conhecimentos incompletos pode escolher. Geralmente, os distratores são respostas plausíveis que correspondem à área de conteúdo.

As perguntas não respondidas são pontuadas como incorretas; não há penalidade por tentar adivinhar. O exame inclui 50 perguntas que afetarão sua pontuação.

Conteúdo não avaliado

O exame inclui 15 perguntas não pontuadas que não afetam sua pontuação. A AWS coleta informações sobre o desempenho do candidato nas perguntas não pontuadas a fim de avaliá-las para uso futuro como perguntas pontuadas. As perguntas não pontuadas não são identificadas no exame.

Resultados do exame

O exame AWS Certified Machine Learning - Specialty (MLS-C01) é um exame de aprovação ou reprovação. O exame é pontuado de acordo com um padrão mínimo estabelecido por profissionais da AWS que seguem as práticas recomendadas e as diretrizes do setor de certificação.

Os resultados do exame são fornecidos como uma pontuação em escala de 100 a 1.000. A pontuação mínima de aprovação é de 750. A pontuação mostra como foi seu desempenho no exame como um todo e se você obteve aprovação ou não. Os modelos de pontuação em escala ajudam a correlacionar as pontuações em várias formas de exame que podem ter níveis de dificuldade ligeiramente diferentes.

O relatório de pontuação pode conter uma tabela de classificação de seu desempenho em cada nível de seção. Essas informações se destinam a fornecer feedback geral sobre seu desempenho no exame. O exame usa um modelo de pontuação compensatória, o que significa que não é necessário obter uma pontuação de aprovação em cada seção. Você precisa passar apenas no exame geral.

Cada seção do exame tem um peso específico, portanto algumas seções têm mais perguntas do que outras. A tabela contém informações gerais que destacam seus pontos fortes e fracos. Tenha cuidado ao interpretar o feedback no nível da seção.

Resumo do conteúdo

Este guia do exame inclui os pesos, os domínios do teste e os objetivos do exame. Não é uma lista abrangente do conteúdo do exame. No entanto, disponibilizamos um contexto adicional para cada um dos objetivos visando ajudar a orientar sua preparação para o exame. A tabela a seguir lista os principais domínios de conteúdo e seus pesos. A tabela precede a descrição completa do conteúdo do exame, que inclui o contexto adicional.

Domínio	% do exame
Domínio 1: Engenharia de dados	20%
Domínio 2: Análise exploratória de dados	24%
Domínio 3: Modelagem	36%
Domínio 4: Implementação e operações de machine learning	20%
TOTAL	100%

Domínio 1: Engenharia de dados

- 1.1 Criar repositórios de dados para machine learning.
 - Identificar a origem dos dados (por exemplo, conteúdo e localização, origens primárias, como dados do usuário)
 - Determinar os meios de armazenamento (por exemplo, banco de dados, data lake, S3, EFS, EBS)
- 1.2 Identificar e implementar uma solução de ingestão de dados.
 - Estilos/tipos de trabalho de dados (carregamento em batch, streaming)
 - Pipelines de ingestão de dados (cargas de trabalho de ML baseadas em batch e cargas de trabalho de ML baseadas em streaming)
 - Kinesis
 - Kinesis Analytics
 - Kinesis Firehose
 - EMR
 - Glue
 - Agendamento de trabalhos
- 1.3 Identificar e implementar uma solução de transformação de dados.
 - Transformar o tráfego de dados (ETL: Glue, EMR, AWS Batch)
 - Manipular dados de ML específicos usando o MapReduce (Hadoop, Spark, Hive)

Domínio 2: Análise exploratória de dados

- 2.1 Limpar e preparar dados para modelagem.
 - Identificar e manipular dados perdidos, dados corrompidos, palavras interrompidas etc.
 - Formatar, normalizar, aumentar e escalar dados
 - Dados rotulados: reconhecer quando você tem dados rotulados suficientes e identificar estratégias de mitigação (ferramentas de rotulação de dados [Mechanical Turk, trabalho manual])

2.2 Executar engenharia de recursos.

- Identificar e extrair recursos de conjuntos de dados, inclusive da origem dos dados, como texto, fala, imagem, conjuntos de dados públicos etc.
- Analisar/avaliar conceitos de engenharia de recursos (agrupamento, tokenização, valores atípicos, recursos sintéticos, transformação de dados, conhecida como “one-hot encoding”, redução da dimensionalidade de dados)

2.3 Analisar e visualizar dados para machine learning.

- Representação gráfica (gráfico de dispersão, série temporal, histograma, diagrama de caixa)
- Interpretação de estatísticas descritivas (correlação, estatísticas resumidas, valor p)
- Agrupamento (hierárquico, diagnóstico, método do cotovelo, tamanho do cluster)

Domínio 3: Modelagem

3.1 Enquadrar problemas de negócios como problemas de machine learning.

- Determinar quando usar ou não usar ML
- Saber a diferença entre aprendizado supervisionado e não supervisionado
- Selecionar entre classificação, regressão, previsão, agrupamento, recomendação etc.

3.2 Selecionar o(s) modelo(s) apropriado(s) para determinado problema de machine learning.

- Xgboost, regressão logística, K-means, regressão linear, árvores de decisão, random forests, RNN, CNN, Ensemble, aprendizado por transferência
- Expressar a intuição por trás de modelos

3.3 Treinar modelos de machine learning.

- Divisão em treino, validação e teste, validação cruzada
- Otimizador, gradiente descendente, funções de perda, ponto mínimo, convergência, lotes, probabilidade etc.
- Escolha de computação (GPU versus CPU, distribuída versus não distribuída, plataforma [Spark versus não-Spark])
- Atualizações e reciclagem de modelos
 - Lote versus tempo real/on-line

3.4 Executar otimização de hiperparâmetros.

- Regularização
 - Dropout
 - L1/L2
- Validação cruzada
- Inicialização do modelo
- Arquitetura de rede neural (camadas/nós), taxa de aprendizado, funções de ativação
- Modelos baseados em árvores (número de árvores, número de níveis)
- Modelos lineares (taxa de aprendizado)

3.5 Avaliar modelos de machine learning.

- Evitar overfitting/underfitting (detecção e manipulação de distorções e variações)
- Métricas (AUC-ROC, exatidão, precisão, recall, RMSE, pontuação F1)
- Matriz de confusão
- Avaliação de modelos off-line e on-line, teste A/B
- Comparação de modelos usando métricas (tempo de treinamento de um modelo, qualidade do modelo, custos de engenharia)
- Validação cruzada

Domínio 4: Implementação e operações de machine learning

4.1 Criar soluções de machine learning para desempenho, disponibilidade, dimensionamento, resiliência e tolerância a falhas.

- Registro e monitoramento do ambiente da AWS
 - CloudTrail e CloudWatch
 - Monitoramento de erros de compilação
- Várias Regiões, várias AZs
- AMI/imagem de referência
- Contêineres do Docker
- Grupos do Auto Scaling
- Dimensionamento correto
 - Instâncias
 - IOPS provisionado
 - Volumes
- Balanceamento de carga
- Práticas recomendadas da AWS

4.2 Recomendar e implementar os serviços e recursos de machine learning apropriados para determinado problema.

- ML na AWS (serviços para aplicativos)
 - Poly
 - Lex
 - Transcribe
- Limites de serviço da AWS
- Criar seu próprio modelo versus algoritmos integrados do SageMaker
- Infraestrutura: (spot, tipos de instância), considerações sobre custo
 - Uso de instâncias spot para treinar modelos de deep learning usando o AWS Batch

4.3 Aplicar práticas básicas de segurança da AWS às soluções de machine learning.

- IAM
- Políticas de bucket do S3
- Grupos de segurança
- VPC
- Criptografia/anonimização

4.4 Implantar e operacionalizar soluções de machine learning.

- Expor endpoints e interagir com eles
- Versionamento de modelo de ML
- Teste A/B
- Treinar pipelines novamente
- Depuração/solução de problemas de ML
 - Detectar e mitigar a queda no desempenho
 - Monitorar o desempenho do modelo

Apêndice

Quais são as principais ferramentas, tecnologias e conceitos que podem ser abordados no exame?

Veja a seguir uma lista (não completa) de ferramentas e tecnologias que podem aparecer no exame. Essa lista está sujeita a alterações e é fornecida para ajudar a entender o escopo geral de serviços, recursos ou tecnologias no exame. As ferramentas e tecnologias gerais dessa lista não aparecem em nenhuma ordem específica. Os produtos da AWS são agrupados de acordo com suas funções principais. Embora algumas dessas tecnologias provavelmente sejam abordadas mais do que outras no exame, a ordem e a colocação delas na lista não são indicativos de importância nem de peso relativo:

- Ingestão/coleta
- Processamento/ETL
- Análise/visualização de dados
- Treinamento do modelo
- Implantação/inferência do modelo
- Operacional
- Serviços para aplicativos de ML da AWS
- Linguagem relevante para ML (por exemplo, Python, Java, Scala, R, SQL)
- Blocos de notas e ambientes de desenvolvimento integrados (IDEs)

Recursos e produtos da AWS

Analytics:

- Amazon Athena
- Amazon EMR
- Amazon Kinesis Data Analytics
- Amazon Kinesis Data Firehose
- Amazon Kinesis Data Streams
- Amazon QuickSight

Computação:

- AWS Batch
- Amazon EC2

Contêineres:

- Amazon Elastic Container Registry (Amazon ECR)
- Amazon Elastic Container Service (Amazon ECS)
- Amazon Elastic Kubernetes Service (Amazon EKS)

Banco de dados:

- AWS Glue
- Amazon Redshift

Internet das Coisas (IoT):

- AWS IoT Greengrass

Machine learning:

- Amazon Comprehend
- AWS Deep Learning AMIs (DLAMI)
- AWS DeepLens
- Amazon Forecast
- Amazon Fraud Detector
- Amazon Lex
- Amazon Polly
- Amazon Rekognition
- Amazon SageMaker
- Amazon Textract
- Amazon Transcribe
- Amazon Translate

Gerenciamento e governança:

- AWS CloudTrail
- Amazon CloudWatch

Redes e entrega de conteúdo:

- Amazon VPC

Segurança, identidade e conformidade:

- AWS Identity and Access Management (IAM)

Sem servidor:

- AWS Fargate
- AWS Lambda

Armazenamento:

- Amazon Elastic File System (Amazon EFS)
- Amazon FSx
- Amazon S3

Recursos e produtos da AWS fora do escopo

Veja a seguir uma lista (não completa) de recursos e produtos da AWS que não são abordados no exame. Estes recursos e produtos não representam todos os serviços da AWS que foram excluídos do conteúdo do exame. Os recursos ou os serviços que não têm nenhuma relação com as funções de trabalho às quais o exame se destina não estão nesta lista porque são considerados irrelevantes.

Os recursos e os serviços da AWS fora do escopo são:

- AWS Data Pipeline
- AWS DeepRacer
- Amazon Machine Learning (Amazon ML)