



Academy-CUR-TF-200-ACMLFO-1-PROD (EN)

Module 01 Student Guide

Version 1.0.4

200-ACMLFO-10-EN-SG

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

This work may not be reproduced or redistributed, in whole or in part,
without prior written permission from Amazon Web Services, Inc.
Commercial copying, lending, or selling is prohibited.

All trademarks are the property of their owners.

Contents

[Module 1: Course Overview](#)

4



Welcome to Module 1: Welcome to AWS Academy Machine Learning.

Module objectives



At the end of this module, you should be able to:

- Identify course prerequisites and objectives
- Describe the various roles that require machine learning knowledge
- Identify resources for further learning
- Explore new technologies at AWS

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

2

At the end of this module, you should be able to:

- Identify course prerequisites and objectives
- Describe the various roles that require machine learning knowledge, and
- Identify resources for further learning
- Explore new technologies at AWS

Module 1: Welcome to AWS Academy Machine Learning Foundations


Section 1: Course prerequisites and objectives

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Introducing Section 1: Course prerequisites and objectives.

Prerequisites



- IT technical knowledge
- Completion of AWS Academy Cloud Foundations
- Intermediate skills with Python programming
- General knowledge of applied statistics

General required knowledge

- Familiarity with cloud computing concepts
- Working knowledge of distributed systems
- Familiarity with general networking concepts
- Working knowledge of multi-tier architectures

Preferred knowledge

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

4

It is helpful to understand the prerequisites of this course.

1. It's preferable for attendees to have some general **IT knowledge**. The foundational computer literacy skills that you need to be successful include basic computer concepts, email, file management, and a good understanding of the internet. You also should have completed the AWS Academy Cloud Foundations course, intermediate skills with Python programming, and general knowledge of applied statistics.
2. **General business knowledge** is important, including insight into how information technology is used in business. Communication skills, leadership abilities, and a customer service orientation are also important skill sets.

To achieve success in this course, you also should have:

- A general familiarity with **cloud computing** concepts
- A working knowledge of **distributed systems**
- Familiarity with **general networking** concepts
- A working knowledge of **multi-tier architectures**

Course objectives



- Describe machine learning (ML)
- Implement a machine learning pipeline using Amazon SageMaker
- Use managed Amazon ML services for forecasting
- Use managed Amazon ML services for computer vision
- Use managed Amazon ML services for natural language processing



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

In this course, you will learn how to describe machine learning (ML), which includes how to:

- Recognize how machine learning and deep learning are part of artificial intelligence
- Describe artificial intelligence and machine learning terminology
- Identify how machine learning can be used to solve a business problem
- Describe the machine learning process
- List the tools available to data scientists
- Identify when to use machine learning instead of traditional software development methods

You will also learn how to implement a machine learning pipeline, which includes learning how to:

- Formulate a problem from a business request
- Obtain and secure data for machine learning
- Use Amazon SageMaker to build a Jupyter notebook
- Outline the process for evaluating data
- Explain why data must be preprocessed
- Use open-source tools to examine and preprocess data
- Use Amazon SageMaker to train and host a machine learning model

- Use cross validation to test the performance of a machine learning model
- Use a hosted model for inference
- Create an Amazon SageMaker hyperparameter tuning job to optimize a model's effectiveness

Finally, you will learn how to use managed Amazon ML services to solve specific machine learning problems in forecasting, computer vision, and natural language processing.

Course outline



Module 1: Welcome to AWS Academy Machine Learning

Module 2: Introducing Machine Learning

Module 3: Implementing a Machine Learning Pipeline with Amazon SageMaker

Module 4: Introducing Forecasting

Module 5: Introducing Computer Vision

Module 6: Introducing Natural Language Processing

Module 7: Course Summary



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

6

To achieve the course objectives, you will complete the following modules:

- Module 2 provides an introduction to machine learning.
- Module 3 describes how to implement a machine learning pipeline with Amazon SageMaker.
- Modules 4, 5, and 6 describe how to apply AWS Managed Services for problems in forecasting, computer vision, and natural language processing.
- Module 7 is a summary of the course, along with an overview of steps that you can take toward attaining certification in machine learning.

The next five slides provide more detail on the subtopics that are covered in each module.

Module 2: Introducing Machine Learning



Module sections:

1. What is machine learning?
2. Business problems that are solved with machine learning
3. Machine learning process
4. Machine learning tools overview
5. Machine learning challenges



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

7

The purpose of this module is to introduce you to major concepts for understanding machine learning.

- Section 1 describes the overall field of machine learning and how machine learning relates to artificial intelligence and deep learning.
- Section 2 summarizes some of the most common business problems that you can solve with machine learning.
- Section 3 describes the general workflow for solving machine learning problems and defines some of the more common terms that are used in machine learning.
- Section 4 reviews some of the tools that machine learning professionals commonly use.
- Section 5 provides an overview of some of the common challenges that you will face when you work with machine learning problems.

Module 3: Implementing a Machine Learning Pipeline with Amazon SageMaker



Module sections:

1. Scenario introduction
2. Collecting and securing data
3. Evaluating your data
4. Feature engineering
5. Training
6. Hosting and using the model
7. Evaluating the accuracy of the model
8. Hyperparameter and model tuning



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

8

Module 3 provides an introduction to Amazon SageMaker and how you can use it to implement a machine learning pipeline. The module focuses on the application of machine learning to computer vision as an example of the machine learning pipeline.

- Section 1 introduces you to the general field of computer vision.
- Sections 2 through 8 describe the phases of the machine learning pipeline by using computer vision as an example application.
 - In Section 2, you learn how to collect and secure data.
 - Section 3 describes different techniques for evaluating data.
 - In Section 4, you learn about the process of feature engineering.
 - Section 5 describes the steps to take in Amazon SageMaker to train a model.
 - Section 6 provides an overview of the options for hosting and using a model with Amazon SageMaker.
 - Finally, Sections 7 and 8 cover how to evaluate and tune your model with Amazon SageMaker.

Module 4: Introducing Forecasting



Module sections:

1. Forecasting overview
2. Processing time series data
3. Using Amazon Forecast



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

9

This module provides an introduction to the use of machine learning to create forecasts that are based on time series data.

- Section 1 describes forecasting in general and summarizes some common applications of forecasting.
- Section 2 outlines some of the pitfalls of the use of time series data to make forecasts.
- Section 3 provides an overview of how to use the Amazon Forecast service.

Module 5: Introducing Computer Vision



Module sections:

1. Introducing computer vision
2. Analyzing image and video
3. Preparing custom datasets for computer vision



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

10

The purpose of this module is to describe how to use machine learning for computer vision.

- Section 1 describes the general problems that must be solved for computer vision.
- Section 2 outlines the process for analyzing images and videos.
- Finally, Section 3 provides the steps to take to prepare datasets for computer vision.

Module 6: Introducing Natural Language Processing



Module sections:

1. Overview of natural language processing (NLP)
2. Natural language processing managed services



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

11

This module provides an introduction to natural language processing with machine learning.

- Section 1 describes the general set of problems that you can solve with natural language processing.
- Section 2 reviews some of the managed AWS services that you can use to address natural language processing problems. The services reviewed include Amazon Transcribe, Amazon Translate, Amazon Lex, Amazon Comprehend, and Amazon Polly.
- Section 3 describes the guided lab in which you will create a bot to schedule appointments.

Module 7: Course Summary



Module sections:

1. Course summary
2. Amazon documentation
3. Certification requirements



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

12

In Module 7, you review what you have learned, and you learn about the process of obtaining a certification with Amazon Machine Learning.

- Section 1 provides an overview of what you have covered in the course.
- In Section 2, you learn more about Amazon documentation and review two common frameworks for applying AWS services.
- Section 3 describes the steps that you should take to continue to work toward an AWS Certified Machine Learning – Specialty certification.

Module 1: Welcome to AWS Academy Machine Learning Foundations

Section 2: Machine learning job roles

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Introducing Section 2: Machine learning job roles.

In this section, you get an overview of some of the more common job roles for machine learning professionals.

Data scientist role



- Applying knowledge of statistics and analytical skills to interpret data
- Data scientists often have degrees in statistics, computer science, or economics
- Some programming skills are required



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

14

If you decide to work toward a data scientist role, you should focus on developing analytical, statistical, and programming skills. As a data scientist, you use those skills to collect, analyze, and interpret large datasets. Some universities now offer degrees in data science, but often data scientists have degrees in statistics, math, computer science, or economics. As a data scientist, you need technical competencies in statistics, machine learning, programming languages, and data analytics.

Machine learning engineer



- Emphasis on programming and system design skills
- Often have background as a developer or software architect
- Some knowledge of statistics required



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

15

If you decide to work toward a career as a machine learning engineer, you need some skills that are similar to a data scientist's skills. However, your focus is more on programming skills and software architecture. As a machine learning engineer, you can apply those skills to design and develop machine learning systems. Machine learning engineers often have previous experience with software development. Although these roles also require technical competencies in statistics and machine learning, they rely more on programming and software engineering than other machine learning roles.

Applied science researcher



- Applies machine learning technology to a specific domain
- Requires knowledge of both the domain and machine learning



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

16

You might also decide to work toward a career in science where you can apply machine learning technology. Machine learning has an impact on everything from astronomy to zoology, so many different paths are open to you. As an applied science researcher, your primary focus is on the type of science that you decide to concentrate on. You need some of the skills of a data scientist, but you must also know how to apply those skills to your chosen domain. These roles also require technical competencies in statistics and machine learning.

Machine learning developer role



- Integrating machine learning with software applications
- Requires strong application development skills and machine learning knowledge



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

17

Many software developers are now integrating machine learning into their applications. If you are working toward a career as a software developer, you should include machine learning technology in your course of study. As a machine learning developer, your primary focus is on software development skills, but you also need some of the skills of a data scientist. Therefore, you should include coursework in statistics and applied mathematics.

Module 1: Welcome to AWS Academy Machine Learning Foundations

Section 3: Resources, documentation, and whitepapers

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Introducing Section 3: Resources, documentation, and whitepapers. This section provides some useful links to extra resources you can use in this course.

Resources and documentation



- [Machine Learning on AWS](#)
- [AWS Machine Learning Blog](#)
- [Machine Learning Solutions in AWS Marketplace](#)
- [Amazon Machine Learning Service Documentation](#)
- [Machine Learning Solutions in AWS Marketplace](#)
- [Amazon Machine Learning Service Documentation](#)
- [Machine Learning in the AWS Partner Network](#)
- [Code Whisperer Installation Instructions](#)
- [Getting Started – Amazon CodeWhisperer](#)
- [Amazon CodeWhisper – Getting Started at AWS Skill Builder](#)
- [Learn how to build an event-driven serverless app](#)

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

19

If you want to learn more about machine learning, you might find the following additional resources helpful:

- [Machine Learning on AWS](#)
- [AWS Machine Learning Blog](#)
- [Machine Learning Solutions in AWS Marketplace](#)
- [Amazon Machine Learning Service Documentation](#)
- [Machine Learning Solutions in AWS Marketplace](#)
- [Amazon Machine Learning Service Documentation](#)
- [Machine Learning in the AWS Partner Network](#)
- [Getting Started – Amazon CodeWhisperer](#)
- [Amazon CodeWhisper – Getting Started at AWS Skill Builder](#)
- [Learn how to build an event-driven serverless app](#)



The “What’s new at AWS” section introduces new and emerging technologies at AWS.

What is Amazon CodeWhisperer?



CodeWhisperer



AI-powered code generator for IDEs and code editors

- **AI coding companion:**
 - Generates code suggestions based on comments and existing code
 - Offers real-time support for code authoring directly within your integrated development environment (IDE)
- **AI security scanner:**
 - Helps identify hard-to-find vulnerabilities
 - References multiple standards and best practices

Content processed by CodeWhisperer Professional is not stored or used for service improvement.

©2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

21

Amazon CodeWhisperer analyzes your comments and code as you write them in your integrated development environment (IDE). It goes beyond code completion by using natural language processing to comprehend the comments in your code. By understanding English comments, CodeWhisperer generates complete functions and code blocks that align with your descriptions. CodeWhisperer also analyzes the surrounding code, ensuring the generated code matches your style and naming conventions and seamlessly integrates into the existing context.

When scanning for security vulnerabilities, CodeWhisperer assesses your code against multiple sets of standards and best practices. This includes the following:

- Open Worldwide Application Security Project (OWASP) standards
- Crypto library best practices
- AWS security standards

The security scan feature is continuously updated to help keep applications free from new security vulnerabilities.

Compatibility: CodeWhisperer integrates with popular tools such as Visual Studio Code, JetBrains IDEs (IntelliJ IDEA, PyCharm, etc.), Amazon SageMaker Studio, JupyterLab, AWS Cloud9, and AWS Lambda console.

Support: CodeWhisperer supports a wide range of programming languages and development environments, including Python, Java, JavaScript, TypeScript, C#, Go, Rust, PHP, Ruby, Kotlin,

C, C++, shell scripting, structured query language (SQL), and Scala.

Installation: You can access CodeWhisperer by downloading and installing the AWS Toolkit IDE extension or plugin. You can also activate CodeWhisperer from directly within the AWS Lambda and AWS Cloud9 console code editors.


Installation instructions vary depending on the environment. For more information, see “Getting started” in the *CodeWhisperer User Guide* at <https://docs.aws.amazon.com/codewhisperer/latest/userguide/getting-started.html>.

Code generation

- Code suggestions
- Code completion
- Code generation from comments
- Alternate code suggestions
- Option to accept or reject
- Reference tracking for code that resembles open-source training data

```
1 import boto3
```

```
1 import boto3
2 # create an s3 bucket named cw95323
```

< 1/3 > Accept tab Accept Word  → ...

```
2 # create an s3 bucket named cw95323
  s3 = boto3.resource('s3')
  s3.create_bucket(Bucket='cw59323')
  # upload a file to the bucket
```



©2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

22

The code generation feature of CodeWhisperer offers code suggestions in real time in your development environment. It automatically offers code completion and code generation suggestions. It uses natural language processing of English comments in your code and an understanding of surrounding code to suggest whole lines of code, complete functions, and logical blocks of code. The generated code is aligned with your coding style and naming conventions. CodeWhisperer prioritizes secure coding and responsible artificial intelligence (responsible AI) practices. It's optimized for Amazon APIs and trained extensively on Amazon and open-source code. You have the option to accept the first suggestion, explore more suggestions, or continue writing your own code. It's important to review each code suggestion before accepting it because you might need to make edits to ensure that the suggestion aligns with your intended functionality.

User actions

- Previous and next suggestion: Use the left arrow and right arrow.
- Accept a suggestion: Press Tab.
- Reject a suggestion: Press Esc.
- Manually start code generation when typing a comment: On MacOS, press Option+C, and on Windows, press Alt+C.

Open Code Reference Log

CodeWhisperer learns from open-source projects and the code it suggests might occasionally resemble code samples from the training data. With the reference log, you can view references to code suggestions that are similar to the training data. When such occurrences happen, CodeWhisperer notifies you and provides repository and licensing information. Use this information to make decisions about whether to use the code in your project and properly attribute the source code as desired.

Security scan

```
1 import boto3
2
3 def upload_file(bucket_name, file_path):
4     s3 = boto3.client('s3')
5     with open(file_path, 'rb') as file:
6         s3.upload_fileobj(file, bucket_name, file_path)
7     print("File uploaded.")
8
9 bucket_name = input("Enter bucket name: ")
10 file_path = input("Enter file path to upload: ")
11 upload_file(bucket_name, file_path)
```

PROBLEMS



CWE-22 – Path traversal: Constructing path names with unsanitized user input can lead to...



©2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

23

The security scanning feature of CodeWhisperer detects security vulnerabilities in both CodeWhisperer-generated code and developer-written code. It scans the code to identify potential vulnerabilities and provides suggestions for remediation. This includes scanning for hard-to-find vulnerabilities that might be overlooked. The security scan is compatible with popular IDEs such as VS Code and JetBrains. It supports Python, Java, and JavaScript.

Benefits of Amazon CodeWhisperer



Value to developers

- Increase velocity.
- Spend less time writing code.
- Receive help directly within your IDE.
- Find security vulnerabilities in your code.



Value to organizations

- Use at all experience levels.
- Support open-source attribution.
- Reduce the risk of security vulnerabilities.
- Increase code quality and developer productivity.



©2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

24

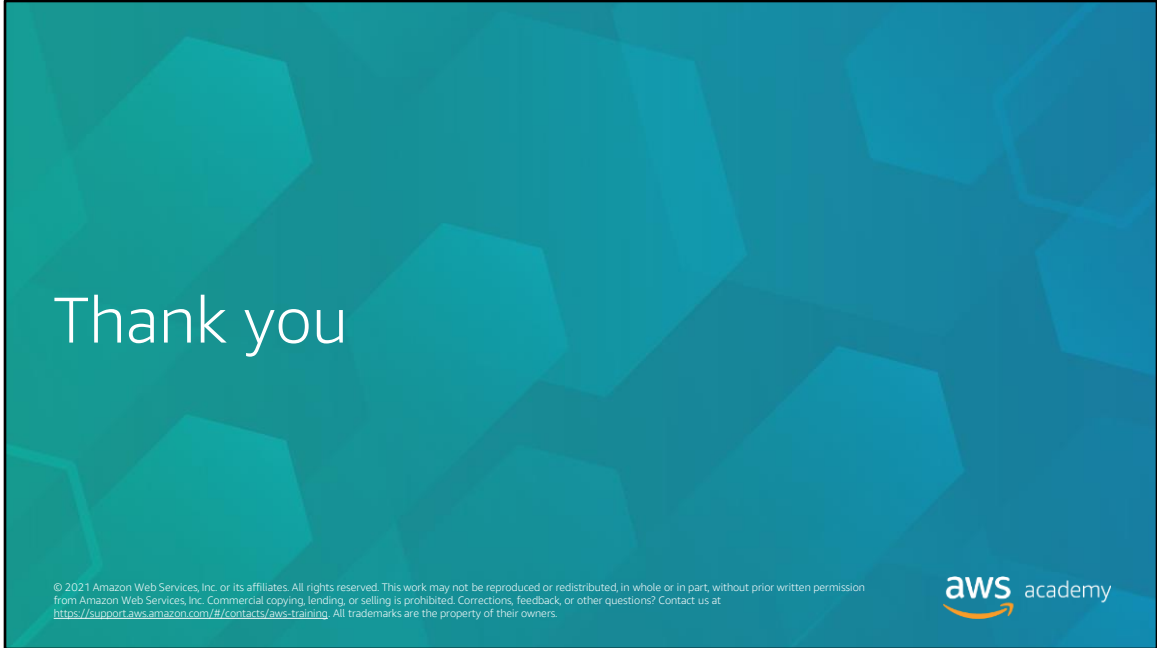
Automated code generation automates repetitive tasks and saves you time. It eliminates the need for you to invest excessive hours in exploring and learning new technologies. Instead, you can rely on high-quality code suggestions that match your coding style. This approach enhances your productivity so you can focus on critical tasks, which encourages innovation and progress in software development. With automated code generation, you can streamline your workflows and achieve significant time savings while ensuring the delivery of code that meets your standards.

CodeWhisperer code generation offers many benefits for software development organizations. It accelerates application development for faster delivery of software solutions. By automating repetitive tasks, it optimizes the use of developer time, so developers can focus on more critical aspects of the project. Additionally, code generation helps mitigate security vulnerabilities, safeguarding the integrity of the codebase.

CodeWhisperer also protects open source intellectual property by providing the open source reference tracker. CodeWhisperer enhances code quality and reliability, leading to robust and efficient applications. And it supports an efficient response to evolving software threats, keeping the codebase up to date with the latest security practices. CodeWhisperer has the potential to increase development speed, security, and the quality of software.

Resources:

- Visit “Getting started” at <https://aws.amazon.com/codewhisperer/resources/>.
- Dive deep with the “Amazon CodeWhisperer – Getting Started” course on AWS Skill Builder at <https://explore.skillbuilder.aws/learn/course/external/view/elearning/16405/amazon-codewhisperer-getting-started>.
- Learn how to build an event-driven serverless app at <https://catalog.us-east-1.prod.workshops.aws/workshops/a33a5d69-1417-4d5f-acc9-ae5c7fba665b/en-US/>.



Thank you for completing this module.