18.0.2 Module 18 Roadmap

Looking Ahead

This week we'll dive deeper into machine learning using unsupervised algorithms, which help us explore data when we're not sure what we're looking for. Now you can analyze data without a clear output in mind.

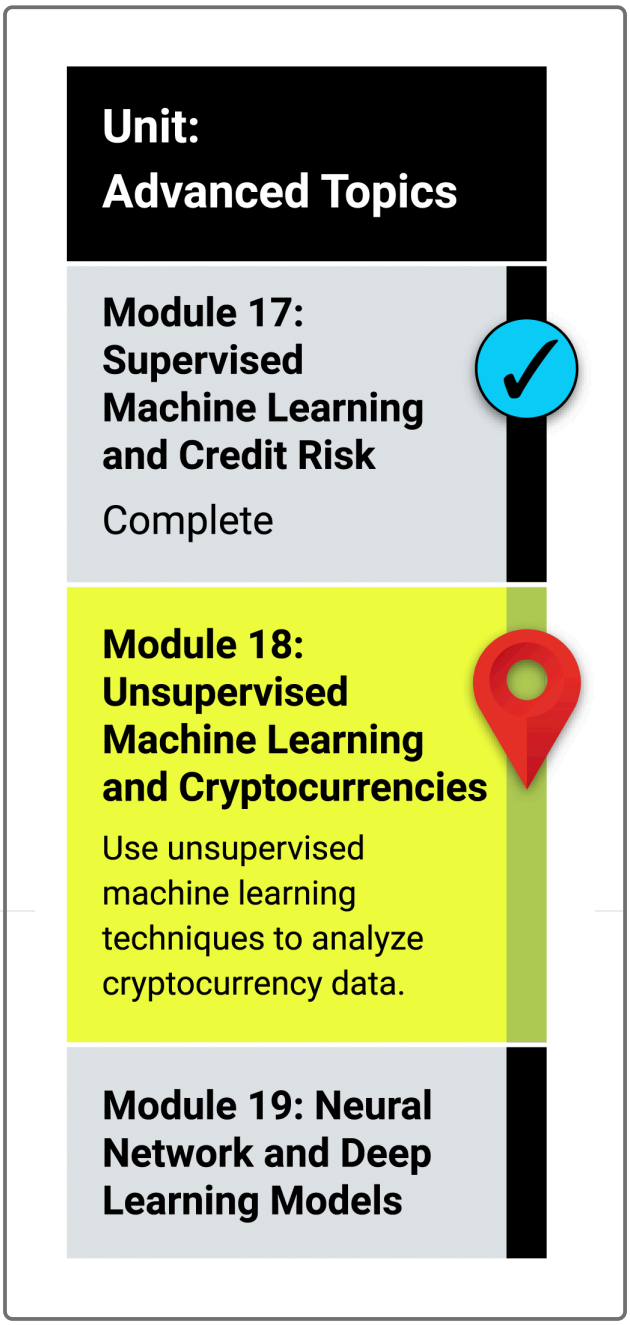
You'll work primarily with the K-means algorithm, the main unsupervised algorithm that groups similar data into clusters. We'll build on this by speeding up the process using principal component analysis (PCA), which employs many different features.

Before starting this module, you should have a strong understanding of training and testing datasets.

What You Will Learn

By the end of this module, you will be able to:

- Describe the differences between supervised and unsupervised learning, including real-world examples of each.
- Preprocess data for unsupervised learning.
- Cluster data using the K-means algorithm.
- Determine the best amount of centroids for K-means using the elbow curve.
- Use PCA to limit features and speed up the model.



Planning Your Schedule

Here's a quick look at the lessons and assignments you'll cover in this module. You can use the time estimates to help pace your learning and plan your schedule:

- Introduction to Module 18 (15 minutes)
- Supervised vs. Unsupervised Learning (1 hour)
- Data Preprocessing (2 hour)
- Clustering and the K-means Algorithm (1 hour)

- Using the Elbow Curve to Find Centroids (1 hour)
- Managing Data Features (2 hour)
- Hierarchical Clustering (2 hour)
- Application (5 hours)