Machine Learning With Big Data

by University of California San Diego

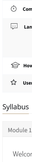
About this Course

Want to make sense of the volumes of data you have collected? Need to incorporate data-driven decisions into your process? This course provides an overview of machine learning techniques to explore, analyze, and leverage data. You will be introduced to tools and algorithms you can use to create machine learning models that learn from data, and to scale those models up to big data problems.

At the end of the course, you will be able to:
- Design an approach to leverage data using the steps in the machine learning process.
- Apply machine learning techniques to explore and prepare data for modeling.
- Identify the type of machine learning problem in order to apply the appropriate set of techniques.
- Construct models that learn from data using widely available open source tools.
- Analyze big data problems using scalable machine learning algorithms on Spark.

Software Requirements:
Cloudera VM, KNIME, Spark

Show less



Taught by:
Mai Nguyen, Lead for Data Analytics
San Diego Supercomputer Center

Taught by:
Ilkay Altintas, Chief Data Science Officer
San Diego Supercomputer Center

Basic Info	Course 4 of 6 in the Big Data Specialization
Commitment	5 Weeks, 3 - 5 hours per week
Language	English, **Subtitles:** Arabic, French, Bengali, Ukrainian, Chinese (Simplified), Greek, Italian, Portuguese (Brazil), Vietnamese, Dutch, Korean, Oriya, German, Pashto, Urdu, Russian, Thai, Indonesian, Swedish, Turkish, Azerbaijani, Spanish, Dari, Hindi, Japanese, Kazakh, Hungarian, Polish
How To Pass	Pass all graded assignments to complete the course.
User Ratings	Average User Rating 4.6

Syllabus

Module 1

Welcome

2 videos

1. **Video:** Welcome to Machine Learning With Big Data
2. **Video:** Summary of Big Data Integration and Processing
3. **Discussion Prompt:** Getting to Know You! Tell us about yourself and why you are taking this course.
4. **Discussion Prompt:** Discussion Forum for Course Content Issues

Show less

Module 2

Introduction to Machine Learning with Big Data

7 videos, 6 readings

1. **Video:** Machine Learning Overview
2. **Video:** Categories Of Machine Learning Techniques
3. **Reading:** Slides: Machine Learning Overview and Applications
4. **Discussion Prompt:** Machine Learning in Everyday Life
5. **Video:** Machine Learning Process
6. **Video:** Goals and Activities in the Machine Learning Process
7. **Video:** CRISP-DM
8. **Video:** Scaling Up Machine Learning Algorithms
9. **Video:** Tools Used in this Course
10. **Reading:** Downloading and Installing Docker Desktop Instructions
11. **Reading:** Introduction to Jupyter Notebooks
12. **Reading:** Downloading Hands-On Materials
13. **Reading:** Basic terminal shell commands
14. **Reading:** Downloading, Installing and Using KNIME

Show less

② **Graded:** Machine Learning Overview

Module 3

Data Exploration

6 videos, 4 readings

1. **Video:** Data Terminology
2. **Video:** Data Exploration
3. **Video:** Data Exploration through Summary Statistics
4. **Video:** Data Exploration through Plots
5. **Discussion Prompt:** What's Wrong with Pie Charts?
6. **Reading:** Slides: Data Exploration Overview and Terminology
7. **Reading:** Description of Daily Weather Dataset
8. **Reading:** Exploring Data with KNIME Plots
9. **Video:** Exploring Data with KNIME Plots
10. **Reading:** Data Exploration in Spark
11. **Video:** Data Exploration in Spark

Show less

② **Graded:** Data Exploration

② **Graded:** Data Exploration in KNIME and Spark Quiz

Module 4

Data Preparation

8 videos, 3 readings

1. **Video:** Data Preparation
2. **Video:** Data Quality
3. **Discussion Prompt:** Quality Issues with Real Data
4. **Video:** Addressing Data Quality Issues
5. **Video:** Feature Selection
6. **Video:** Feature Transformation
7. **Video:** Dimensionality Reduction
8. **Discussion Prompt:** Domain Knowledge in Data Preparation
9. **Reading:** Slides: Data Preparation for Machine Learning
10. **Reading:** Handling Missing Values in KNIME
11. **Video:** Handling Missing Values in KNIME
12. **Reading:** Handling Missing Values in Spark
13. **Video:** Handling Missing Values in Spark

Show less

② **Graded:** Data Preparation

② **Graded:** Handling Missing Values in KNIME and Spark Quiz

Module 5

Classification

8 videos, 5 readings

1. **Video:** Classification
2. **Video:** Building and Applying a Classification Model
3. **Reading:** Slides: What is Classification?
4. **Video:** Classification Algorithms
5. **Video:** k-Nearest Neighbors
6. **Video:** Decision Trees
7. **Video:** Naïve Bayes
8. **Reading:** Slides: Classification Algorithms
9. **Reading:** Classification using Decision Tree in KNIME
10. **Video:** Classification using Decision Tree in KNIME
11. **Reading:** Interpreting a Decision Tree in KNIME
12. **Reading:** Classification in Spark
13. **Video:** Classification in Spark
14. **Discussion Prompt:** Why Exclude Relative Humidity?

Show less

② **Graded:** Classification

② **Graded:** Classification in KNIME and Spark Quiz

Module 6

Evaluation of Machine Learning Models

7 videos, 6 readings

1. **Video:** Generalization and Overfitting
2. **Video:** Overfitting in Decision Trees
3. **Video:** Using a Validation Set
4. **Reading:** Slides: Overfitting: What is it and how would you prevent it?
5. **Video:** Metrics to Evaluate Model Performance
6. **Video:** Confusion Matrix
7. **Discussion Prompt:** Model Interpretability vs. Accuracy
8. **Reading:** Slides: Model evaluation metrics and methods
9. **Reading:** Evaluation of Decision Tree in KNIME
10. **Video:** Evaluation of Decision Tree in KNIME
11. **Reading:** Completed KNIME Workflows
12. **Reading:** Evaluation of Decision Tree in Spark
13. **Video:** Evaluation of Decision Tree in Spark
14. **Reading:** Comparing Classification Results for KNIME and Spark

Show less

② **Graded:** Model Evaluation

② **Graded:** Model Evaluation in KNIME and Spark Quiz

Module 7

Regression, Cluster Analysis, and Association Analysis

8 videos, 5 readings

1. **Video:** Regression Overview
2. **Video:** Linear Regression
3. **Reading:** Slides: Regression
4. **Video:** Cluster Analysis
5. **Video:** k-Means Clustering
6. **Reading:** Slides: Cluster Analysis
7. **Discussion Prompt:** Clustering Applications
8. **Video:** Association Analysis
9. **Video:** Association Analysis in Detail
10. **Reading:** Slides: Association Analysis
11. **Discussion Prompt:** Applications of Association Analysis
12. **Video:** Machine Learning With Big Data - Final Remarks
13. **Reading:** Description of Minute Weather Dataset
14. **Reading:** Cluster Analysis in Spark
15. **Video:** Cluster Analysis in Spark

Show less

② **Graded:** Regression, Cluster Analysis, & Association Analysis

② **Graded:** Cluster Analysis in Spark Quiz

View Less

How It Works

General

What do start dates and end dates mean?

Once you enroll,
More

Course 4 of Specialization

Unlock Value in Massive Datasets

Learn fundamental big data methods in six straightforward courses.

Big Data
University of California San Diego

Learn More

View the course in catalog

Related Courses

Big Data - Capstone Project
University of California San Diego

Introduction to Big Data
University of California San Diego

Graph Analytics for Big Data
University of California San Diego

Big Data Integration and Processing
University of California San Diego

Big Data Modeling and Management Systems
University of California San Diego