

CS M10ML: CLOUD DATA SCIENCE AND MACHINE LEARNING

Originator

egarcia

Co-Contributor(s)
Name(s)

Nikjeh, Esmaail (enikjeh)

Loay, Alnaji (lalnaji)

College

Moorpark College

Attach Support Documentation (as needed)

The importance of Data Science with Cloud Computing.pdf
 Meeting_12_16_2020.docx
 Cloud Computing Regional Advisory Minutes 5-29-2020.pdf
 Spark Data Engineer - Contractor - United States - Indeed.pdf
 Data Engineer, Machine Learning - Remote - Indeed.pdf
 Business Analyst - Chatsworth, CA 91311 - Indeed.pdf
 Splunk_consutant.pdf
 Analytics Jobs, Employment in Ventura County, CA _ Indeed.pdf
 Scorpion - Senior Data Scientist.pdf
 EtiVenture Inc Jobs and Careers _ Indeed.pdf
 Cloud Data Science Jobs, Employment in Ventura County, CA _ Indeed.pdf
 CS M10ML_state approval letter_CCC000624063.pdf

Discipline (CB01A)

CS - Computer Science

Course Number (CB01B)

M10ML

Course Title (CB02)

Cloud Data Science and Machine Learning

Banner/Short Title

Cloud Data Sci Machine Learn

Credit Type

Credit

Start Term

Fall 2022

Catalog Course Description

Utilizes various cloud based Data Science tools and services available from Amazon Web Services (AWS) cloud platform to perform data science analysis and machine learning. Uses various examples of how data science and machine learning is used to process vast amounts of collected data to derive predictive analysis. Uses tools such as AWS SageMaker and Amazon Forecast and computer vision tools such as Amazon Rekognition, Amazon Ground Truth, image and video processing and other AWS data science services. Uses Natural Language processing such as Amazon Polly, Amazon Comprehend, and Amazon Translate. Uses AWS Academy Data Science and Machine Learning curriculum. Aligns to Data Science industry professional certification training. Prepares students to pass relevant AWS data science professional certification.

Taxonomy of Programs (TOP) Code (CB03)

0707.10 - *Computer Programming

Course Credit Status (CB04)

D (Credit - Degree Applicable)

Course Transfer Status (CB05) (select one only)

A (Transferable to both UC and CSU)

Course Basic Skills Status (CB08)

N - The Course is Not a Basic Skills Course

SAM Priority Code (CB09)

D - Possibly Occupational

Course Cooperative Work Experience Education Status (CB10)

N - Is Not Part of a Cooperative Work Experience Education Program

Course Classification Status (CB11)

Y - Credit Course

Educational Assistance Class Instruction (Approved Special Class) (CB13)

N - The Course is Not an Approved Special Class

Course Prior to Transfer Level (CB21)

Y - Not Applicable

Course Noncredit Category (CB22)

Y - Credit Course

Funding Agency Category (CB23)

Y - Not Applicable (Funding Not Used)

Course Program Status (CB24)

1 - Program Applicable

General Education Status (CB25)

Y - Not Applicable

Support Course Status (CB26)

N - Course is not a support course

Field trips

Will not be required

Grading method

(L) Letter Graded

Alternate grading methods

(O) Student Option- Letter/Pass

(E) Credit by exam, license, etc.

(P) Pass/No Pass Grading

Does this course require an instructional materials fee?

No

Repeatable for Credit

No

Is this course part of a family?

No

Units and Hours

Carnegie Unit Override

No

In-Class**Lecture****Minimum Contact/In-Class Lecture Hours**

17.5

Maximum Contact/In-Class Lecture Hours

17.5

Activity**Laboratory****Minimum Contact/In-Class Laboratory Hours**

52.5

Maximum Contact/In-Class Laboratory Hours

52.5

Total in-Class**Total in-Class****Total Minimum Contact/In-Class Hours**

87.5

Total Maximum Contact/In-Class Hours

87.5

Outside-of-Class**Internship/Cooperative Work Experience****Paid****Unpaid****Total Outside-of-Class****Total Outside-of-Class****Minimum Outside-of-Class Hours**

35

Maximum Outside-of-Class Hours

35

Total Student Learning**Total Student Learning****Total Minimum Student Learning Hours**

157.5

Total Maximum Student Learning Hours

157.5

Minimum Units (CB07)

2

Maximum Units (CB06)

2

Prerequisites

MATH M15 or MATH M15H and CS M10DS

Advisories on Recommended Preparation

CS M10R and CS M10P

Entrance Skills

Entrance Skills

MATH M15 or MATH M15H, and CS M10DS

Prerequisite Course Objectives

CS M10DS-deploy basic statistical concepts that data scientists need to know like measure of central tendency, percentiles, probability distributions, dimensionality reduction, over and under sampling, and Bayesian statistics.

CS M10DS-distinguish fundamental aspects of machine learning algorithms.

CS M10DS-frame problems to enable suitable solutions via machine learning.

CS M10DS-train (process of modeling) and evaluate machine learning models.

CS M10DS-deploy machine learning models into operations.

CS M10DS-build prediction, categorization and recommendation APIs (application program interface).

CS M10DS-deploy tools for collaborative and social programming.

CS M10DS-generate high-quality graphical and textual results.

CS M10DS-find similar items, pattern discovery (Neural Network and Machine Learning), mining data streams, frequent itemsets, link analysis, and mining graph data.

CS M10DS-distinguish supervised machine learning algorithms like k nearest neighbors, decision trees, Naive Bayes, regression, and support vector machines.

CS M10DS-distinguish unsupervised machine learning (data mining) algorithms like classification, k-means clustering, evaluation of clustering, hierarchical clustering, spectral (partitional) clustering, and Neural networks.

MATH M15-summarize data graphically by displaying data using methods from descriptive statistics, interpreting data in tables graphically by using histograms, frequency distributions, box-and whisker plots (five-number summary); find measures of central tendency for data sets: mean, median, and mode; find measures of variation for data sets: standard deviation, variance, and range; determine relative positions of data and distinguish among scales of measurements and their implications; distinguish between populations and samples; and identify the standard method of obtaining data and the advantages and disadvantages of each.

MATH M15-find simple probabilities and probabilities of compound events and compute probabilities using the complement, discrete probability distributions; apply concepts of sample space, and the binomial probability distribution.

MATH M15-standardize a normally distributed random variable; use normal distribution tables to find probabilities for normally distributed random variables and the t-distribution; use the Central Limit Theorem to find probabilities for sampling distributions.

MATH M15-construct and interpret confidence intervals for proportions and means.

MATH M15-identify the basics of hypothesis testing and perform hypothesis testing for means, proportions and standard deviations from one population, and difference of means and proportions from two populations, including finding and interpreting p-value and examining Type I and Type II error.

MATH M15-find linear least-squares regression equations for appropriate data sets; graph least-square regression equations on the scatter plot for the data sets; find and apply the coefficient of correlation.

MATH M15-use the chi-square distribution to test independence and to test goodness of fit.

MATH M15-conduct a one-way Analysis of Variance (ANOVA) hypothesis test.

MATH M15-select an appropriate hypothesis test and interpret the result using p-value; use appropriate statistical technique to analyze and interpret applications based on data related to business, social sciences, psychology, life sciences, health sciences or education, and interpret results using technology-based statistical analysis.

MATH M15H-summarize data graphically by displaying data using methods from descriptive statistics, interpreting data in tables graphically by using histograms, frequency distributions, box-and whisker plots (five-number summary); find measures of central tendency for data sets: mean, median, and mode; find measures of variation for data sets: standard deviation, variance, and range; determine relative positions of data and distinguish among scales of measurements and their implications; distinguish between populations and samples; and identify the standard method of obtaining data and the advantages and disadvantages of each.

MATH M15H-find simple probabilities and probabilities of compound events and compute probabilities using the complement, discrete probability distributions; apply concepts of sample space, and the binomial probability distribution. Honors: And the Poisson distribution.

MATH M15H-standardize a normally distributed random variable; use normal distribution tables to find probabilities for normally distributed random variables and the t-distribution; use the Central Limit Theorem to find probabilities for sampling distributions.

MATH M15H-construct and interpret confidence intervals for proportions and means.

MATH M15H-identify the basics of hypothesis testing and perform hypothesis testing for means, proportions and standard deviations from one population, and difference of means and proportions from two populations, including finding and interpreting p-value and examining Type I and Type II error.

MATH M15H-find linear least-squares regression equations for appropriate data sets; graph least-square regression equations on the scatter plot for the data sets; find and apply the coefficient of correlation.

MATH M15H-use the chi-square distribution to test independence and to test goodness of fit.

MATH M15H-conduct a one-way Analysis of Variance (ANOVA) hypothesis test.

MATH M15H-select an appropriate hypothesis test and interpret the result using p-value; use appropriate statistical technique to analyze and interpret applications based on data related to business, social sciences, psychology, life sciences, health sciences or education, and interpret results using technology-based statistical analysis.

MATH M15H-Honors: find probabilities using Poisson distributions.

MATH M15H-Honors: estimate binomial probabilities.

MATH M15H-Honors: find and use standard deviations.

Requisite Justification**Requisite Type**

Prerequisite

Requisite

MATH M15 or MATH M15H and CS M10DS

Requisite Description

Course not in a sequence

Level of Scrutiny/Justification

Closely related lecture/laboratory course

Requisite Type

Recommended Preparation

Requisite

CS M10R and CS M10P

Requisite Description

Course not in a sequence

Level of Scrutiny/Justification

Content review

Student Learning Outcomes (CSLOs)**Upon satisfactory completion of the course, students will be able to:**

- | | |
|---|---|
| 1 | apply cloud data science services that solve data analysis inquiries. |
| 2 | use a cloud based Jupyter notebook and other cloud tools and services to extract, transform and load the needed data analytics processes. |

Course Objectives**Upon satisfactory completion of the course, students will be able to:**

- | | |
|---|--|
| 1 | relate cloud infrastructure of services to big data problems. |
| 2 | analyze data with cloud services. |
| 3 | describe cloud infrastructure needed to manage and analyze big data. |
| 4 | manage, secure and scale compute and storage instances on the cloud. |
| 5 | identify analytical cloud services. |
| 6 | configure and manage storage options. |

Course Content**Lecture/Course Content****25% - Big Data Analytical concepts:**

- Ingest, store, and secure data
- Query a data store with manual schema specification
- Query a data store with automated schema generation
- Load and query data in a data warehouse
- Visualize structured and unstructured data
- Automate loading data into a data warehouse
- Analyze unstructured data
- Analyze IoT data

10% - Machine Learning:

- Business problems
- Learning process and challenges
- Learning tools

15% - Using Sagemaker for Machine Learning:

- Collecting and securing data
- Evaluating your data
- Feature engineering
- Training a Model Using Amazon SageMaker
- Hyperparameter and model tuning

10% - Forecasting:

- Processing time series data
- Using Amazon Forecast

20% - Computer Vision:

- Analyzing image and video
- Introducing Amazon Rekognition
- Preparing custom datasets for computer vision
- Labeling images with Amazon Ground Truth
- Facial Recognition

20% - Natural Language Processing:

- Natural Language Processing managed services
- Introducing Amazon Polly
- Introducing Amazon Comprehend
- Introducing Amazon Translate
- Using Bot to schedule appointments

Laboratory or Activity Content

- 5% - Lab 1 - Ingesting Data into Amazon S3
- 5% - Lab 2 - Querying Amazon S3 Data Using Amazon Athena
- 10% - Lab 3 - Transforming Data Using Amazon S3, AWS Glue, and Amazon Athena
- 5% - Lab 4 - Loading the Amazon Redshift Cluster with data and querying
- 10% - Lab 5 - Analyze data with Amazon SageMaker, Jupyter Notebooks, and Bokeh
- 5% - Lab 6 - Executing a data pipeline to load data into Amazon S3
- 15% - Lab 7 - Streaming data with AWS Kinesis Firehose, Amazon Elasticsearch Service, and Kibana
- 5% - Lab 8 - Using AWS IoT analytics for data ingestion and analysis
- 20% - Lab 9 - Using Sagemaker to create, import, explore, encode, train with a data pipeline and apply hyperparameter tuning
- 5% - Lab 10 - Time Series and Amazon Forecasting
- 5% - Lab 11 - Computer Vision and Facial Recognition
- 10% - Lab 12 - Natural Language processing and creating a chatbot

Methods of Evaluation

Which of these methods will students use to demonstrate proficiency in the subject matter of this course? (Check all that apply):

Written expression
 Problem solving exercises
 Skills demonstrations

Methods of Evaluation may include, but are not limited to, the following typical classroom assessment techniques/required assignments (check as many as are deemed appropriate):

Computational homework
 Group projects
 Individual projects
 Laboratory activities
 Oral analysis/critiques
 Oral presentations
 Problem-solving exams
 Quizzes
 Reports/papers

Research papers
 Simulations
 Skills demonstrations
 Skills tests or practical examinations
 Written homework

Instructional Methodology

Specify the methods of instruction that may be employed in this course

Case studies
 Class activities
 Class discussions
 Collaborative group work
 Computer-aided presentations
 Distance Education
 Group discussions
 Guest speakers
 Instructor-guided interpretation and analysis
 Instructor-guided use of technology
 Internet research
 Laboratory activities
 Large group activities
 Lecture
 Observation
 Small group activities
 Web-based presentations

Describe specific examples of the methods the instructor will use:

The instructor will explain and demonstrate how cloud services can be used to solve data science related problems by the following teaching techniques:

- demonstrate various cloud based services and how they are used to collect, process and visualize the data collected and processed for decision making.
- explain various python and R program libraries and programming code used in data science.
- explain and interpret statistic method being used and how it relates to a specific data science problem.
- demonstrate various visualization methodologies and criteria used to select the most appropriate visual display.

Representative Course Assignments

Writing Assignments

- Writing an explanation that utilizes Amazon SageMaker notebook instance to create a Jupyter notebook, explain programming code and the use of markdown cells within the notebook.
- Documenting encoded categorical variables using data repositories such as the automobile dataset from the UC Irvine Machine Learning Repository.

Critical Thinking Assignments

- Analyzing test data to make a prediction or forecast of future data performance. Example: Compare the forecasted data with actual data to determine accuracy of prediction.
- Interpreting a confusion matrix that explains and supports decisions based on forecasting results. Example: Interpret visual chart and how specific chart supports decision making.

Reading Assignments

- Reading a Python and R program code that when executed produces a correct result, then analyze and interpret the program code. Example; Given a python code, explain the code and use of Panda library code to generate output.
- Reviewing Amazon Web Services to gain familiarity with Sagemaker and cloud services related to data science. Example: Within AWS console, explain the various data science and machine learning services available and their most appropriate application.

Skills Demonstrations

- Analyzing debugged output to produce the needed programming results. Example: Capture debug code output verifying correct use of programming code such boolean variables, output variable values, and use of verbose coding techniques.

- Applying visualization techniques that best support needed decision making of data science computational results. For Example: display output data in 2 different visual formats and explain how one view is superior or a better fit in supporting a data science decision.

Outside Assignments

Representative Outside Assignments

- Writing Python and R programs to solve data science problems such as using pre-written panda libraries to prepare a time series, then use Amazon Forecast to create a predictor, and forecasted data and then compare the forecast data against actual known data. Example: Comparison of forecasted versus actual known data will reveal accuracy of prediction model. This is done by inputting 90% of actual data, run forecasting algorithm to predict next set of data, then take 10% of retained real data and compare this to predicted data to determine accuracy of prediction model.
- Creating a custom image collection for Amazon Rekognition, and detect known faces in an image. For example: Review a Youtube video where Facial Recognition is used to allow/deny a pet entering a home. The facial recognition uses software and a camera to make the decision to allow the pet cat entry into the home only when it has no prey (ie: bird/mouse/hunted game) in its mouth.

Articulation

Equivalent Courses at other CCCs

College	Course ID	Course Title	Units
Santa Monica	CS 79F	Machine Learning of AWS	3

District General Education

A. Natural Sciences

B. Social and Behavioral Sciences

C. Humanities

D. Language and Rationality

E. Health and Physical Education/Kinesiology

F. Ethnic Studies/Gender Studies

Course is CSU transferable

Yes

CSU Baccalaureate List effective term:

F2021

CSU GE-Breadth**Area A: English Language Communication and Critical Thinking****Area B: Scientific Inquiry and Quantitative Reasoning****Area C: Arts and Humanities****Area D: Social Sciences****Area E: Lifelong Learning and Self-Development****Area F: Ethnic Studies****CSU Graduation Requirement in U.S. History, Constitution and American Ideals:****UC TCA****UC TCA**
Approved**Effective term:**
Fall 2021**IGETC****Area 1: English Communication****Area 2A: Mathematical Concepts & Quantitative Reasoning****Area 3: Arts and Humanities****Area 4: Social and Behavioral Sciences****Area 5: Physical and Biological Sciences****Area 6: Languages Other than English (LOTE)****Textbooks and Lab Manuals****Resource Type**
Other Resource Type**Description**

Students will be enrolled in Amazon Web Services Online Academy course(s) in Data Science and Machine Learning which provides the most recent Technology Training resources which includes lectures, demos, quizzes, and free labs up to \$100 of cloud based computational credits per student for free. Students will be able to use "State of the Art" modern technical tools used by top professionals in the Data Science Industry. AWS currently has a 70% market share in Data Science and companies moving to cloud are discovering performance benefits of up to 90% improvement due to super computing power only available on cloud platforms. Additionally, new innovations data prediction algorithms, forecasting models and new Data Science services are being introduced into AWS data management tool-sets. The AWS Academy provide access to the most modern tools and advancements in the Data Science market where students can explore new innovations. Advancements in this area have been tremendous and evolving and would not be appropriately taught using a traditional textbook since some topics could be no longer relevant. New solutions, services, and techniques are being introduced and are reviewed at AWS website as part of the course learning. Students in AWS Academy are provided access to production based real world cloud services that solve modern data science problems using the latest cloud

Resource Type
Textbook

Classic Textbook

No

Description

Barth, Antje and Chris Fregly. *Data Science on AWS: Implementing End-to-End, Continuous AI and Machine Learning Pipelines*. O'Reilly Media, 2021.

Resource Type

Textbook

Classic Textbook

Yes

Description

Singh, Himanshu. *Practical Machine Learning with AWS: Process, Build, Deploy, and Productionize Your Models Using AWS*. Apress, 2020.

Library Resources**Assignments requiring library resources**

Use of Safari O'Reilly Technical database subscription for students. Research statistical models, python and R programming language, and data science ebooks.

Sufficient Library Resources exist

Yes

Example of Assignments Requiring Library Resources

Interpreting a confusion matrix that explains and supports decisions based on forecasting results. Example: Interpret visual chart and how specific chart supports decision making.

Distance Education Addendum**Definitions****Distance Education Modalities**

Hybrid (1%–50% online)

Hybrid (51%–99% online)

100% online

Faculty Certifications

Faculty assigned to teach Hybrid or Fully Online sections of this course will receive training in how to satisfy the Federal and state regulations governing regular effective/substantive contact for distance education. The training will include common elements in the district-supported learning management system (LMS), online teaching methods, regular effective/substantive contact, and best practices.

Yes

Faculty assigned to teach Hybrid or Fully Online sections of this course will meet with the EAC Alternate Media Specialist to ensure that the course content meets the required Federal and state accessibility standards for access by students with disabilities. Common areas for discussion include accessibility of PDF files, images, captioning of videos, Power Point presentations, math and scientific notation, and ensuring the use of style mark-up in Word documents.

Yes

Regular Effective/Substantive Contact**Hybrid (1%–50% online) Modality:****Method of Instruction**

Asynchronous Dialog (e.g., discussion board)

Document typical activities or assignments for each method of instruction

Instructor will post a question or an erroneous python or R code, students will respond to the question or attempt to analyze the code and suggest how to debug and fix the code.

E-mail	Instructor will email students with announcements about the course or an upcoming event. Students in turn may email the instructor with their questions or concerns. Students will email their programs and projects to the instructor.
Face to Face (by student request; cannot be required)	Students will have the option to meet the instructor and work in the computer lab in the presence of the instructor to get one-on-one help from the instructor.
Other DE (e.g., recorded lectures)	Instructor may record the lectures and post them for students to view within a specified time frame to be ready for the accompanying data science assignments. Students will upload their programs and projects to the course webpage.
Synchronous Dialog (e.g., online chat)	Instructor may be available on a certain day or days of the week within a certain time frame to help students and answer their questions via an online chat.
Telephone	Instructor may provide a phone number for the students where they can leave a voicemail and expect a call back within 24 hours.
Video Conferencing	Instructor may be available on a certain day or days of the week within a certain time frame to help students and answer their questions via live video conferencing.
Hybrid (51%–99% online) Modality:	
Method of Instruction	Document typical activities or assignments for each method of instruction
Asynchronous Dialog (e.g., discussion board)	Instructor will post a question or an erroneous python or R code, students will respond to the question or attempt to analyze the code and suggest how to debug and fix the code.
E-mail	Instructor will email students with announcements about the course or an upcoming event. Students in turn may email the instructor with their questions or concerns. Students will email their programs and projects to the instructor.
Face to Face (by student request; cannot be required)	Students will have the option to meet the instructor and work in the computer lab in the presence of the instructor to get one-on-one help from the instructor.
Other DE (e.g., recorded lectures)	Instructor may record the lectures and post them for students to view within a specified time frame to be ready for the accompanying programming assignments. Students will upload their programs and projects to the course webpage.
Synchronous Dialog (e.g., online chat)	Instructor may be available on a certain day or days of the week within a certain time frame to help students and answer their questions via an online chat.
Telephone	Instructor may provide a phone number for the students where they can leave a voicemail and expect a call back within 24 hours.
Video Conferencing	Instructor may be available on a certain day or days of the week within a certain time frame to help students and answer their questions via live video conferencing.
100% online Modality:	
Method of Instruction	Document typical activities or assignments for each method of instruction
Asynchronous Dialog (e.g., discussion board)	Instructor will post a question or an erroneous python or R code, students will respond to the question or attempt to analyze the code and suggest how to debug and fix the code.
E-mail	Instructor will email students with announcements about the course or an upcoming event. Students in turn may email the instructor with their questions or concerns. Students will email their programs and projects to the instructor.
Face to Face (by student request; cannot be required)	Students will have the option to meet the instructor and work in the computer lab in the presence of the instructor to get one-on-one help from the instructor.

Other DE (e.g., recorded lectures)	Instructor may record the lectures and post them for students to view within a specified time frame to be ready for the accompanying programming assignments. Students will upload their programs and projects to the course webpage.
Synchronous Dialog (e.g., online chat)	Instructor may be available on a certain day or days of the week within a certain time frame to help students and answer their questions via an online chat.
Telephone	Instructor may provide a phone number for the students where they can leave a voicemail and expect a call back within 24 hours.
Video Conferencing	Instructor may be available on a certain day or days of the week within a certain time frame to help students and answer their questions via live video conferencing.

Examinations

Hybrid (1%–50% online) Modality

On campus
Online

Hybrid (51%–99% online) Modality

On campus
Online

Primary Minimum Qualification

COMPUTER SCIENCE

Additional local certifications required

An approved Amazon Web Services (AWS) Academy Instructor which requires: (about 120 hours)

- 1 - demonstrated competency in AWS cloud computing services.
- 2 - successfully completing curriculum of course being taught, including all chapter exams and final exam at 90%, and completion of all AWS labs.
- 3 - successfully passing an AWS Professional Certification Exam.
- 4 - successfully passing an online Technical Interview demonstrating mastery in a variety of course related topics.
- 5 - completion and continuous maintenance of AWS Academy duties and responsibilities.
- 6 - maintain active status as an approved AWS Academy Instructor.

Review and Approval Dates

Department Chair

02/15/2022

Dean

02/15/2022

Technical Review

02/17/2022

Curriculum Committee

3/1/2022

DTRW-I

MM/DD/YYYY

Curriculum Committee

MM/DD/YYYY

Board

MM/DD/YYYY

CCCCO

MM/DD/YYYY

Control Number

CCC000624063

DOE/accreditation approval date

MM/DD/YYYY