

Certified Professional Machine Learning & Data Science Foundation (CP-MLDS Foundation) Certification Course



<http://cpmlsf.devopsppalliance.org/>
<http://devopsppalliance.org/cp-mlsf.html>



info@devopsppalliance.org



[/devopspp](https://twitter.com/devopspp)



[/devopspp](https://www.facebook.com/devopspp)

<http://devopsppalliance.org/>



What is CP-MLDS Foundation course?

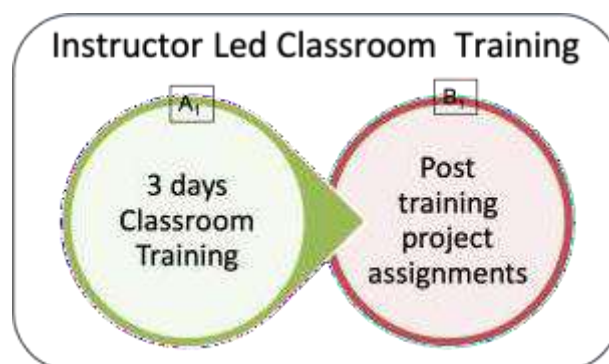
CP-MLDS Foundation stands for “Certified Professional – Machine Learning and Data Science Foundation” certification prepared and honored by “Agile Testing Alliance”.

The course is applicable for all roles and knowledge, experience & certification is consciously designed for all those who want to learning practical Machine learning and Data Science.

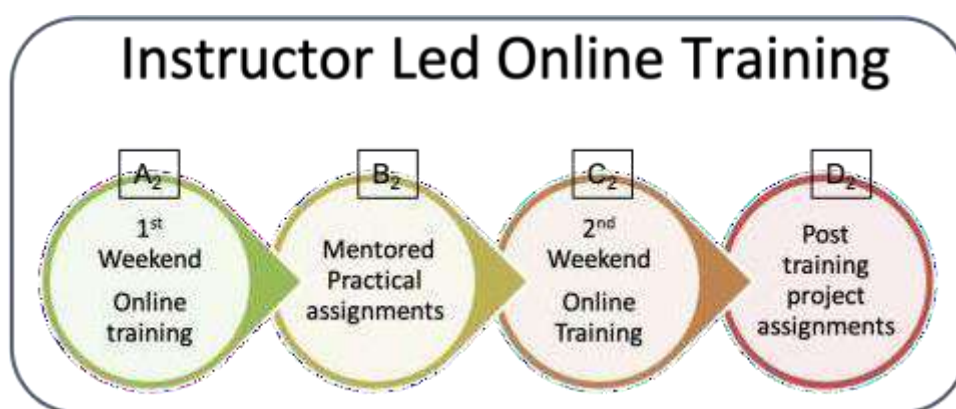
What is new in this version 1.3.1 of CP-MLDS

As per this latest version of learning objectives,

- a. Module 4 for Linear Regression now includes Regularization Techniques of
 - i. Lasso Regression and
 - ii. Ridge Regression
- b. Module 5 for classification is added with Decision Tree based classification technique.
- c. Enhancement in CP-MLDS Learning Structure by introducing Online Training Mode, which would help participants to join the training program anywhere across the Globe –
 - I. Instructor Led Classroom Training Mode - At the same time, the learning structure has been modified to include mentored practical assignments post the 3 days training.



- II. Instructor Led Online Training Mode – Introduction of Instructor LED Online training Mode to the CP-MLDS Learning Structure to have 4 half days Online sessions over two consecutive weekends. Intermediate Week is meant for mentored practical assignments.



How is CP-MLDS useful?

Machine learning is based on algorithms that can learn from data without relying on rules-based programming.

As ever more of the analog world gets digitized, our ability to learn from data by developing and testing algorithms will only become more important for what are now seen as traditional businesses. Google chief economist Hal Varian calls this “computer kaizen.”

For “just as mass production changed the way products were assembled and continuous improvement changed how manufacturing was done,” he says, “so continuous [and often automatic] experimentation will improve the way we optimize business processes in our organizations.

This is where it is clear, that we are into a computer kaizen world. A world where Machine learning and Data science algorithms are driving this self-learning and continuous learning process and bringing about a massive change in almost every industry. We have machine learning and data science now being used in banking, insurance, health care, sports, manufacturing, smart cities, IoT Solutioning, Automotive, Aviation, Shipping and every other industry for that matter.

Machine learning and Data science need thus has increased multifold in past few years and would keep on increasing. At the same time there is a dearth of experienced professionals who know Machine learning and Data Science. There is no dearth of machine learning and data science programs where folks would need to spend infinite amount of time and efforts to acquire this knowledge. The challenge is for working professionals to spend so much time. Most often this rigor is lost over few weeks or months.

This program solves this issue addresses two basic needs

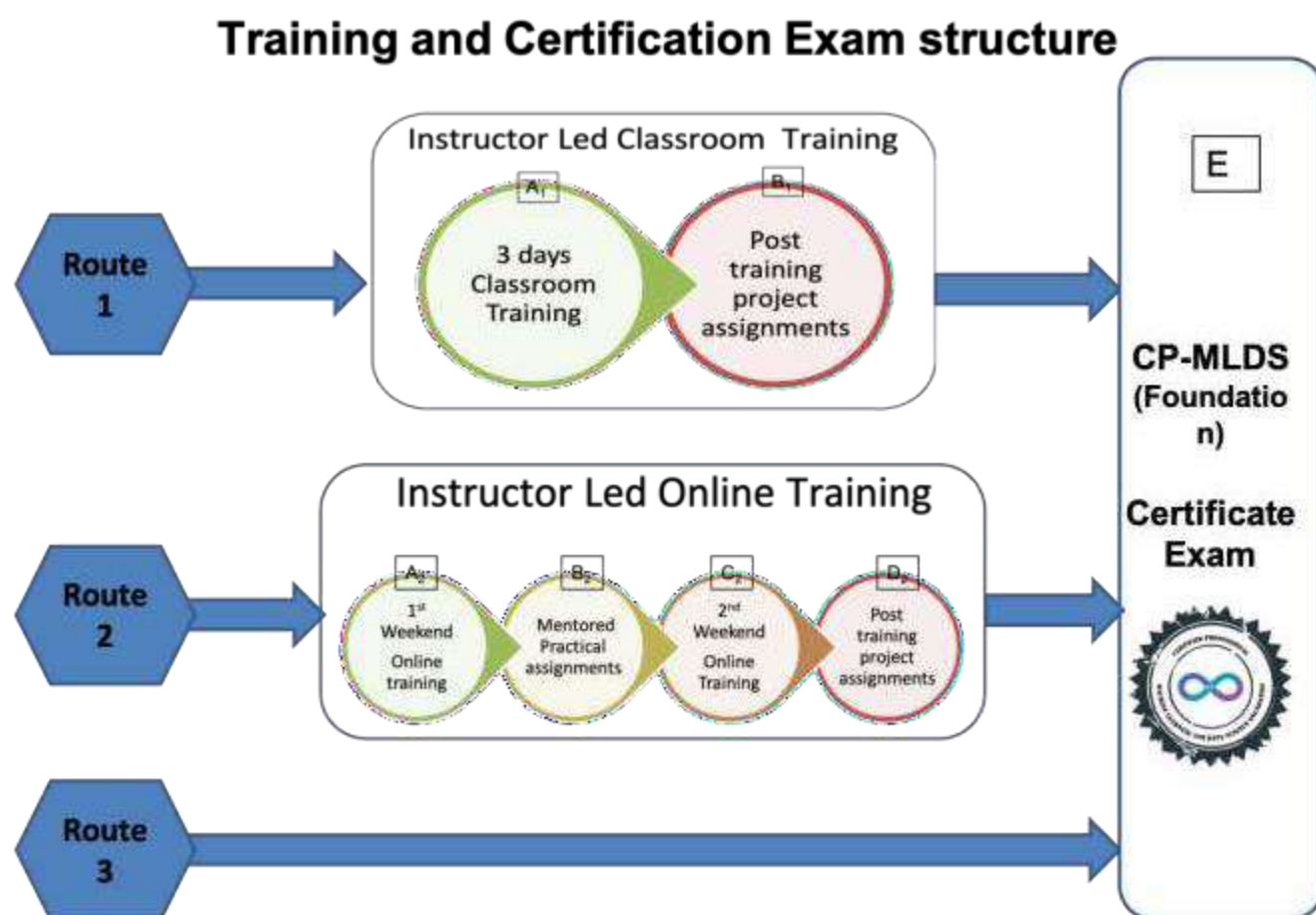
- a) Practical tool-based Machine learning and Data Science exposure for every working professional
- b) Allow working professionals to acquire this knowledge in the most agile manner

Am I Eligible?

There are no pre-requisites for this certification program except having some prior knowledge of any programming language and basics of mathematics and statistics. Program is Python driven and having prior knowledge of Python would be an advantage.

What is the Training and Certification Exam structure?

CP-MLDS is one of the only **FastTrack** program specially designed to Reskill / Upskill working professionals.



The Training and Certification structure is composed of
A₁ – Instructor Led Classroom sessions – 3 full days (24 Hrs) Instructor led fully hands on sessions

B₁ - Project assignments and Mentoring – Post the three days training program, 2 projects assignments will be shared by the mentor. The process is designed to help the participants strengthen their learnings.

A₂ – Instructor Led Online sessions-Part1 – Over the 1st Weekend (5 Hrs each day) Instructor led Online sessions

B₂ - Project assignments and Mentoring – Post the 1st weekend Online training sessions, 2 projects assignments will be shared by the mentor. The process is designed to help the participants strengthen their learnings.

C₂ – Instructor Led Online sessions-Part2 – Over the 2nd Weekend (5 Hrs each day) Instructor led Online sessions

D₂ - Project assignments and Mentoring – Post the online training program, 2 projects assignments will be shared by the mentor. The process is designed to help the participants strengthen their learnings.

E - Certification Exam. The certification exam comprises of one Theory Section (40 Marks) and one Practical section (60 Marks). Theory section is of 1 hour and practical section is of 2 hours. Getting 60% in both the sections of exam is necessary to get the CP-MLDS certificate.

- The training program (Part [A₁+B₁] or Part [A₂+B₂+C₂+D₂] above) is not mandatory for someone to go for CP-MLDS certification.
- Thus certification exam can be taken by following either of the three routes –
 - **Route 1** - by going through a formal Lab Training (step [A₁+B₁] above) and then going for the certification exam.
 - **Route 2** - by going through a formal Instructor Led Online Training (step [A₂+B₂+C₂+D₂] above) and then going for the certification exam E.
 - **Route 3** – **This is a direct route** which allows someone to appear for the exam E and get the CPMLDS (Foundation) certification after passing the exam. This route is recommended for someone already familiar with the subject.

Learning Objectives and Main Topics

Module 1: The World of AI – ML (Machine Learning)

Topics:

- A. Introduction to AI
- B. AI (Artificial Intelligence), ML (Machine Learning) and DL(Deep Learning)
- C. What is Machine Learning?
- D. Applications of Machine Learning in various industries
- E. Types of Machine Learning
 - Supervised Learning
 - Unsupervised Learning
- F. Maths refresher
 - Advanced Stats and probability
 - Calculus and Algebra

Learning Objectives:

At the end of this module students should

- Understand the types of Machine Learning and the names of algorithms used
- Be able to tell the real-world applications of Machine Learning
- Be able to answer questions related to Linear Algebra, Calculus and Statistics

Module 2: Setting up Environment and Recap of Python for Machine Learning

Topics:

- A. Setup for Python and JetBrains' Pycharm (IDE)
- B. Installing libraries required for Machine Learning
- C. Introduction to Jupyter notebook
- D. Recap Python
 - Data types
 - Operators
 - Conditions and Loops

- Data structures – List, Dictionary, Set and Tuples
 - functions
- E. Introduction to libraries of Python useful for ML
- NumPy – Numerical processing
 - Pandas – Data Processing
 - Matplotlib, Seaborn – Data Visualizations in Python

Learning Objectives:

At the end of this module students should

- Setup an environment in Anaconda for Machine Learning
- Independently install various libraries required for Machine Learning
- Solve small programming assignments in python using NumPy, Pandas,
- Matplotlib and Seaborn Libraries.

Module 3: EDA - Exploratory Data Analysis

Topics:

- A. Types of data
- B. Using summary statistics to understand data
- C. Using boxplot to visualize data
- D. Pre-processing Data
- Handling missing data and outliers
 - Normalization
- E. Visualizing correlations between features

Learning Objectives:

At the end of this module students should

- Understand data that is used for Machine Learning
- Understand and identify missing data, outliers
- Pre-processing data – Data imputation, normalization
- Explain summary statistics for a data frame
- Understand correlation between columns

Module 4: Simple and Multiple Linear Regression and Regularization techniques

Topics:

- A. Understanding Regression problems
- B. What is Simple and Multiple Linear Regression
- C. Understand maths and relevant concepts behind the Linear Regression
 - Understanding the linear equation Features
 - Learning a model for $y=wx+b$ equation and parameters
 - Defining and visualizing the cost function
 - Understanding Gradient Descent algorithm for cost optimization
- D. Performance Metrics for Linear Regression
- E. Hands on Exercises Using SciKit-learn library for Machine Learning for Simple and Multiple Linear Regression datasets
- F. Regularization Techniques
 - Lasso Regression
 - Ridge Regression

Learning Objectives:

At the end of this module participants

- Will understand basic concepts and maths of Simple and Multiple Linear Regression.
- Will be able to use scikit-learn library for applying linear regression
- Compute Performance metrics of the Linear model
- Understand Regularisation techniques using Lasso Regression and Ridge Regression to overcome overfitting

Module 5: Classification using Logistic Regression (LR) and Decision Trees

Topics:

- A. Understand Classification problems and differentiate it from regression problems.
- B. Understanding maths and relevant concepts for Logistic regression
 - The Sigmoid function

- Cost function for classification
 - Finding derivative of sigmoid function
- C. Understanding Decision Tree technique for classification
- D. Hands on Exercises for classification problem with Logistic Regression and Decision Tree using Scikit-learn library
- E. Performance Measures for Classification
- Confusion Matrix
 - Precision
 - Recall

Learning Objectives:

At the end of this module students

- Understands and easily differentiate between Regression problems and classification problems
- Will be able to apply logistic regression and Decision Tree to different datasets using scikit-learn
- Would be able to determine performance metrics for the given problems

Module 6: Introduction to Clustering

Topics:

- A. Understanding UnSupervised ML problem
- B. Understand Clustering technique
- C. Types of Clustering
- D. Applications of Clustering
- E. Understanding k-means clustering algorithm
- F. Learn to implement clustering using scikit-learn

Learning Objectives:

At the end of this module students

- Understands what is clustering and its application areas
- Implementing clustering using scikit-learn to different datasets

Module 7: Introduction to NLP (Natural Language Processing)

Topics:

- A. Understanding NLP
- B. Application areas of NLP
- G. Understand preprocessing and terminology used for NLP using NLTK library

Learning Objectives:

At the end of this module students

- Understands what is NLP wrt Text Data and its application areas
- Preprocessing the Text Data with NLTK library