Caltech | Center for Technology & Management Education

California Institute of Technology
Machine Learning for Advanced Analytics
2022–2023
Syllabus
===

Course Page: https://ctme.caltech.edu/ml-open

Course Description
Analytics is understanding the semantics of data which allows us to find meaningful patterns and knowledge in such data. Analytics predicts the future based on analysis of the data. Machines learn from data using analytics concepts.

Machine Learning applications are found in almost all industries. The success of today's enterprises depends largely on their workers' ability to make better and faster decisions, ones that are based on facts/data to solve complicated business problems. Machine Learning has become the key enabler for such a decision-making philosophy—commonly referred to as "evidence-based management."

Machine Learning is a subfield of computer science that gives "computers the ability to learn without being explicitly programmed." Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, Machine Learning explores the study and construction of algorithms that can learn from data and make predictions about future events. Machine Learning modeling methods will be covered in this course using R and Python (Scikit-Learn and TensorFlow) statistical packages.

This content of the course has 3 sections. The first part covers Regression where the predicted variable is numerical is nature. The second part covers classification where the predicted variable is categorical in nature. And the third part covers non-supervisory learning methods like clustering. Deep Learning concepts will also be introduced on the third day of the course.

The following modeling methods will be covered – kNN, Regression, Logistic Regression, Decision Trees, Naïve Bayes, Support Vector Machines and Neural Networks.

This course will cover the underlying mathematical concepts for a wide variety of Machine Learning methods and algorithms, plus various procedures used to assess the value and validity of them. Course participants will apply Machine Learning tools (R and Python) to build predictive models.
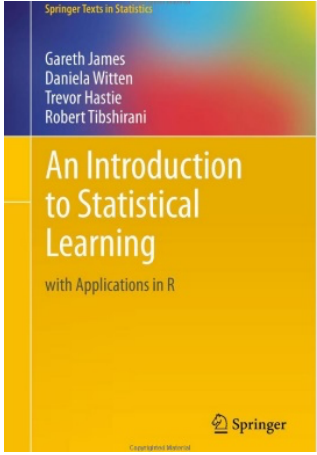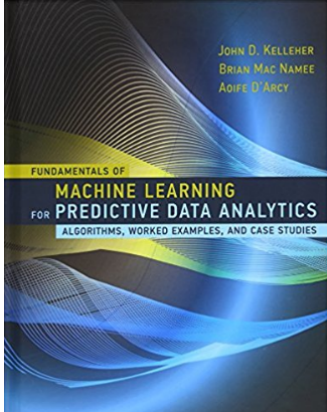
Specific topics include:

> **Modeling Methods**
> - Linear Regression
> - k Nearest Neighbor
> - Logistic Regression
> - Support Vector Machine
> - Decision Trees
> - Naïve Bayes
> - Clustering
> - Neural Networks + Deep Learning
>
> **Modeling Tools**
> - R and Python (Scikit-Learn + TensorFlow)
>   Both R and Python are open-source platforms which can be downloaded at no cost. Currently Python platforms for machine learning - Scikit-Learn and TensorFlow are heavily promoted by Google.

Text Books

1	**An Introduction to Statistical Learning With Applications in R (ISLR)** By • Gareth James • Daniela Witten • Trevor Hastie • Robert Tibshirani Publisher: Springer	
2	**Fundamentals of Machine Learning for Predictive Data Analytics** By • John D. Kelleher • Brian Mac Namee • Aoife D'Arcy Publisher: MIT Press Optional	

Caltech | **Center for Technology & Management Education**

Lesson Plan

	Theme	Lesson Content
1	Tools for ML + Regression	• Machine Learning and Deep Learning • Tools for ML: R and Python/Colab + GCP • Linear Regression: Introduction
2	Regression	• Regression Assessment • Gradient Descent • Regression Inference + Multi Var Reg
3	Regression	• Regression Categorical Variables • Regression Quality
4	Classification	• Logistic Regression • Naïve Bayes • kNN Model - k Nearest Neighbor
5	Classification	• Naïve Bayes • Shannon's Entropy Model + Decision Trees • Support Vector Machine
6	Unsupervised Learning	• Clustering • Principal Component Analysis

Caltech | Center for Technology & Management Education

Course Objectives

At the end of this course, participants will be able to:

- Understand the role of ML (Machine Learning) in the broader spectrum of Analytics.
- Know the range of methods and algorithms used for machine learning.
- Apply different ML methods and algorithms to practical business problems/data.
- Have the hands-on experience to build and assess variety of ML models using a well-known ML software tool – R and Python (Scikit-Learn + TensorFlow).
- Compare different model developed by different algorithms using a set of assessment methods.
- Can deploy ML models developed by different types of methods and algorithms.
- Recommend the best machine learning algorithms for detecting trends in large, noisy data sets.
- Assess machine learning algorithms for accuracy using standard techniques such as cross-validation.
- Categorize data mining and machine learning techniques based on their mathematical assumptions, scalability, limitations, and parameters.

Instructor Information

Dr. Ash Pahwa
Office Phone: (949) 378-1229
Email: ash@ashpahwa.com
Website: www.AshPahwa.com

Ash Pahwa, Ph.D., is an educator, author, entrepreneur, and technology visionary with three decades of industry and academic experience. He has founded several successful technology companies during his career, the latest of which is A+ Web Services.

Dr. Pahwa earned his doctorate in Computer Science from the Illinois Institute of Technology in Chicago. He is listed in *Who's Who in the Frontiers of Science and Technology*. He is also a Google Certified Analytics Consultant. His expertise includes Machine Learning, Deep Learning, Natural Language Processing, Digital Image Processing, and data storage technologies. He has worked with many fortune 500 companies on Machine/Deep Learning projects. He has published many papers in Machine Learning/Predicted Analytics conferences.

In Industry, Dr. Pahwa has worked for General Electric, AT&T Bell Laboratories, Xerox Corporation, and Oracle. He founded CD-Gen, Inc. and DV Studio Technologies, LLC., which introduced successful products for CD-Recording (CDR) and MPEG encoding. His book, *CD-Recordable Bible* was published in English, Japanese, and German.

In Academia, Dr. Pahwa teaches courses at California Institute of Technology (Pasadena) and the University of California system. Since 2008, he taught many courses at UC Irvine, UCLA, and UC San Diego.