# Advanced Web Analytics:
# Harnessing the Predictive Power

## Syllabus
### UCSD Extension

**Class Meeting Information**

Start Date: February 1, 2016
End Date: March 14, 2016
7 class meetings
Online

**Instructor Information**

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

**Ash Pahwa, Ph.D.**

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 search engine optimization, web analytics, web programming, digital image processing, database management, digital video, and data storage technologies.

**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 internet technology courses and conducts webinars in the University of California system. Since 2008, he taught many courses at UC Irvine, UCLA, and UC San Diego, including:

| Website Development | Digital Marketing | Predictive Analytics |
|---|---|---|
| WordPress CMS | Search Engine Optimization | R Programming |
| Microsoft ASP.NET | Web Analytics | Predictive Analytics Using Google Analytics |
| | Google AdWords | MATLAB Programming |

**Prerequisites — Classes or Knowledge Required for this Course**

- Basic Math: Statistics, Functions, Matrix
- Basic Programming
- Google Analytics

## Course Description

Predictive Analytics (PA) is a leading-edge technology that is being adopted by Fortune 500 corporations and coveted by many other entities in industry and academia. As the name suggests, it seeks to predict the outcome of certain events.

Google Analytics (GA) provides sophisticated traffic information about a website and it delivers a comprehensive array of business intelligence and visitor behavior insights. GA data can be used to predict future events. The goal of this course is to teach how to effectively use GA data by building predictive models.

One of the most difficult things in PA is the acquisition of high quality input data. If the input data is inaccurate or incomplete, predictive modeling results will be equally flawed. Since GA data is always complete and accurate, this resource can be used very effectively to predict some future events or trends.

R is a scripting language for statistical data manipulation and analysis. One of the primary applications of the R package is PA. Like Google Analytics, it is freely available. Once the predictive models are constructed, GA data is fed into R which becomes the work engine of the analysis. Beside R there are many tools commercially available for PA. KNIME is one of the open source software which can be used. Spreadsheet software (like Microsoft Excel) can be used for modeling verification and validation.

This course is configured for professionals who are currently working with Google Analytics. It will provide insight into how the GA data can be used for prediction of future events using R package. This course will first cover the Google Analytics, Predictive Analytics techniques, and the R Statistical package. It will also discuss the problems faced in predicting events and trends.

This course will focus on the following 4 predictive analytics techniques – Linear Regression, Logistic Regression, Naïve Bayes, and Decision Trees. These techniques will be discussed in detail using the Google Analytics data.

In the end these techniques will be compared with each other to show which technique works best for which application.

## Learning Objectives

The goal of this course is to provide immersion into the predictive power of Google Analytics. Students will understand the following:

- Understand Google Analytics as the source of data for Predictive Analytics
- Review all Predictive Analytics techniques
- Understand R statistical package
- Interface to Google Analytics from R
- Analyze problems real world analyst face in collecting data for Predictive Analytics

**Methodology:** This course will be taught online.

**Course Text**
None

## Evaluation and Grading

### Evaluation of Student Performance

| | |
|---|---|
| Home work assignments 10 points each (7) | 70 points |
| Total | 70 points |

### Grading Scale

A = 90% – 100%  B = 80% – 89%  C = 70% – 79%  D = 60% – 69%

## Course Outline

| Lesson | **Subject** |
|---|---|
| 1 | Introduction to Analytics<br>• What is Predictive Analytics? CRISP/DM model<br>• Data Problems in Predictive Analytics |
| 2 | Google Analytics (GA)<br>• Page Views + Visits + Visitor + Time<br>• Unique Visitors<br>• Traffic Sources + Content Metrics + Link Tagging |
| 3 | Tools for PA:<br>• Excel<br>• R<br>• KNIME |
| 4 | Linear Regression<br>• 2 Variables<br>• Multi-variable |
| 5 | Logistic Regression<br>• 2 Variables<br>• Multi Variable |
| 6 | Naïve Bayes<br>• Naïve Bayes Mathematics<br>• PA Using Naïve Bayes |
| 7 | Decision Trees<br>• Decision Tree Math<br>• Decision Tree Using Excel + KNIME + R |