

Syllabus and Learning Guide

Course Title: Data Analytics Powered by Google

Number: IDH 3034/ IDH 3035

Schedule: Fall/Spring Mondays 9:00AM – 11:45AM

Term: Fall/Spring 2018

Credits: 3

Pre-requisites: Honors College Standing

Instructor Information

- Professor: Miguel Alonso Jr
- Department: School of Computing and Information Sciences
- Office location/hours: Mondays 3pm – 5pm, ECS 254A
- Telephone: 305-741-6441
- Email: malonsoj@fiu.edu
- Website: <http://dsail.fiu.edu>

Why should you care about Data Analytics?

How does Google know about your specific account details when you call customer service? How does American Express know which customers will close their accounts in the next quarter? How does Capital One know what deals to offer its customers and when? The answer is Big Data and Data Analytics. Data analytics (DA) is the processing of large data sets with the goal of extracting insights about the information they contain, using software and systems. These techniques and technology are used in commercial industries to enable organizations to make more-informed business decisions. They are also used by scientists and researchers to verify or disprove scientific models, theories and hypotheses. Throughout this two-term course, I'll be your guide through the exciting world of big data and data analytics. And best of all, you will be able to answer, "How does Google know what I want to buy?"

How will this course help you succeed?

This course introduces data analytics using the free tools available online from Google. The course may make use of Google personnel and Google clients as instructors. Students will be eligible for internships at Google and Google clients. The goal of the course is to develop a basic understanding on the theory and practice of machine learning from a variety of perspectives and to gain the fundamentals on applying machine learning techniques to real-world applications.

By the end of the course, you will be able to answer the following questions:

- What is "data analytics"?
- What is the cloud, cloud infrastructure, and how is it deployed and managed?
- What is machine learning and how is it deployed in the cloud?

- What are the ethical implications of deploying data analytics and machine learning systems?
- What is bias in Machine Learning and how can I design systems that mitigate these biases?
- What are the Google Infrastructure and Analytics tools that I can use to create and execute a data analytics project?

Course Topics

Here is a tentative topic outline of the major topics that will be covered in the course.

Semester 1: Cloud Infrastructure and Tools

1. Introduction, Google Cloud Platform Fundamental: Core Infrastructure
2. Architecting with Google Cloud Platform: Infrastructure
3. Introduction to Cloud Identity
4. Architecting with Google Cloud Platform: Design and Process
5. Ethics and Bias in Data and Machine Learning

Semester 2: Data Analytics and Machine Learning Tools

1. From Data to Insights with GCP
2. Google Cloud Platform Fundamentals: Big Data and Machine Learning
3. Data Engineering on Google Cloud Platform
4. Machine Learning with TensorFlow on Google Cloud Platform
5. Global socio-economic impacts of Big Data, Data Mining, and Machine Learning

Where can you look for specific information?

Although much of the topics, algorithms, and information that will be discussed is readily available online, I recommend the following books, as they will unquestionably serve as an invaluable resource throughout your career in machine learning.

Google Cloud Platform for Architects: Design and Manage Powerful Cloud Solutions, by Janani Ravi, Judy Raj, and Vitthal Srinivasan

Hands-On Machine Learning on Google Cloud Platform: Implementing Smart and Efficient Analytics Using Cloud ML Engine

Additionally, there will be a multitude of resources, including videos, tutorials, and additional information on the canvas course shell. Please refer to the canvas course shell for additional information on course updates, changes, and general communications.

How will you succeed in this course?

Students need to have a foundational understanding of computers, and a basic understanding of statistics and algebra. Additionally, knowing a programming language such as python and having a general understanding of linear algebra, although not required, will be helpful. That is, students entering the class with basic knowledge of probability, statistics and algorithms will be

at an advantage, but the class will be designed so that anyone with basic mathematical background can catch up and fully participate.

Keep in mind, that education, the process of creating knowledge, values and information, is going through a transformation from me being the 'sage on stage' towards student-centered learning. Your learning is in your hands. Research shows that knowledge is better created, retained and used by students when acquired through active, participatory learning rather than being a passive recipient (Fink 2003). Understanding is not downloadable. Deep, meaningful learning (as opposed to memorization) is fostered and cultivated by you engaging with the material. You will learn in proportion to the amount of effort you put into it.

How will you and I evaluate your progress?

Final Team Presentation (50%) – A final team presentation for each term (the first project will be about cloud infrastructure and the second will be about data analytics and machine learning) will be made in which you and your team will present your final project, complete with project report and demonstration, to a panel of judges.

Mid-term Team Presentation (20%) – In order to make sure that the final project is moving forward in a timely fashion, you and your team will present the current state of the project approximately mid-way through each semester.

Weekly Labs (20%) Since this is an applied course in cloud infrastructure and data analytics, there will not be any exams. However, you must complete weekly or bi-weekly labs in order to make progress through learning Google's tools. The labs will be started during class and completed either in class or at home.

Class participation (10%) Showing up is half the battle! Again, since this is an applied course, class participation is of utmost importance. You should be showing up to each and every class and furthermore, you should be participating in all of the in-class and out-of-class activities.

Your final grade will be assigned based on the following grading scheme:

A: 95% or above	C+: 77% - 79%
A- : 90% – 94%	C: 70% - 76%
B+: 87% – 89%	D: 60% - 69%
B: 83% – 86%	F: 59% or less
B-: 80% – 82%	

What will we do in this course?

We will be learning about cloud computing, big data, and machine learning and applying them towards solving hard problems. You will learn to use Google's Cloud Platform towards designing and implementing a data analytics solution using large datasets and deploying data analytics projects in the cloud.

Academic Integrity at FIU

FIU CORE Values: Responsibility, Truth, Freedom, Respect & Excellence

- *All students are expected to adhere to a standard of academic conduct, which demonstrates respect for themselves, their fellow students, and the core values.*
- *All students should understand that if they are found responsible for academic misconduct, they will be subject to the FIU Academic Misconduct Policies & Procedures. The FIU Academic Integrity home page (<http://integrity.fiu.edu>) provides a flow chart that illustrates the informal and formal resolution process. The Informal Resolution Form is available for completion online.*

I abide and endorse the FIU's policy on academic integrity. Any form of academic misconduct is considered a serious offense. Should you have academic or personal problems that are getting in the way of your academic success, please contact your instructor.

FIU's Policy for academic misconduct includes these definitions for these intentional acts or omissions:

Cheating: The unauthorized use of books, notes, aids, electronic sources; or unauthorized use of on-line exams, library materials or assistance from another person with respect to examinations, course assignments, field service reports, class recitations; or the unauthorized possession of examination papers (or on-line examination) or course materials, whether originally authorized or not. Any student helping another cheat may be found guilty of academic misconduct

Plagiarism: The deliberate use and appropriation of another's work without any indication of the source and the representation of such work as the student's own. Any student, who fails to give credit for ideas, expressions or materials taken from another source, including internet sources, is guilty of plagiarism. Any student helping another to plagiarize may be found guilty of academic misconduct.

Self-Plagiarism: This is using your own work for another assignment without providing a citation indicating that this work was previously used. When citing yourself, use cite in-text citations to identify yourself as the author.

Misrepresentation: Intentionally lying to a member of the faculty, staff, administration, or an outside agency to gain academic advantage for oneself or another, or to misrepresent or in other ways interfere with the investigation of a charge of academic misconduct.

Misuse of Computer Services: The unauthorized use of any computer, computer resource or computer project number, or the alteration or destruction of computerized information or files or unauthorized appropriation of another's program(s).

Bribery: The offering of money or any item or service to a member of the faculty, staff, administration or any other person in order to commit academic misconduct.

Conspiracy and Collusion: The planning or acting with one or more fellow students, any member of the faculty, staff or administration, or any other person to commit any form of academic misconduct together.

Falsification of Records: The tampering with or altering in any way any academic record used or maintained by the University.

Academic Dishonesty: In general, by any act or omission not specifically mentioned above and which is outside the customary scope of preparing and completing academic assignments and/or contrary to the above stated policies concerning academic integrity.

More questions?

I am happy to help you either in my office or via email. I have offices hours, however, if you would like to meet outside of those hours, please send me your request via email and I will do my best to accommodate.

If you send me an email, please adhere to the following guidelines:

1. Always have a subject line that includes IDH3034/IDH3035.
2. It should have a proper greeting.
3. Please sign your name.

****As we work together through this learning experience, I will assess the needs of the class and may update the syllabus accordingly. Therefore, the syllabus is subject to change****

Course Learning Objectives

By the end of the course, the student will:

- Develop an understanding of basic concepts in data analytics and related cloud infrastructure
- Develop an understanding of basic concepts in machine learning
- Develop an understanding of the ethical implications of data analytics and machine learning in government, the private sector, and globally
- Develop an understanding of bias in machine learning algorithms and strategies to mitigate the effects of bias in ML applications
- Develop an understanding of how to use Google tools online to create and execute a data analytics project

Course Learning Outcomes

By the end of the course(s), the student will develop:

- An ability to apply knowledge of data analytics to solve complex problems

- An ability to clean data, select algorithms and execute a successful data analysis
- An awareness of the ethical issues that arise when applying data analytics and machine learning to real-world problems
- An appreciation for the implications of bias in machine learning and data analytics applications
- A skill set sufficient to qualify for an internship with a client of Google or Google