COURSE SHORT DESCRIPTION

In this course we'll explore how a range of companies are using big data to achieve business advantage. Designed to enhance and extend the content in the core "Business Analytics" course, we will use a combination of case studies and hands-on projects to give you experience in identifying opportunities for leveraging data to support, make, and implement decisions. We will also provide an overview of the key engineering tools used for working with big data, providing an introduction to some of the core projects used for both batch and real time data processing such as Cassandra, Hadoop, Kafka and Storm.

COURSE LEARNING OBJECTIVES

1. **Additional applications of data science** - we'll look at a wider range of problems and data science based solutions than were covered in the core "Business Analytics" course.

2. **Additional algorithms** - we'll look at a number of additional algorithms for extracting knowledge and changing behavior based on data.

3. **Engineering big data** - we'll provide an introduction to some of the core projects used for both batch and real time data processing such as Cassandra, Hadoop, Kafka and Storm.
COURSE ROADMAP/SCHEDULE

Session 1. Identifying Opportunities - 6 hours
Case studies and content looking for opportunities to leverage data to provide business advantage. The case studies will include:

- Improving conversion funnels using A/B tests
- Recommendation engines for e-commerce
- Models for fraud detection in online payment systems
- Sentiment analysis from social media
- Analyzing images

Session 2. Selecting Algorithms - 6 hours
Based on the problems identified in the first session, we'll look at algorithms we can use to solve the problems and how to select between algorithms for various classes of problem. We’ll also look at how MapReduce, graph theory, Bayesian probability and neural networks can be used for extracting knowledge out of data. We will cover:

- Practical algorithms for classification, regression and clustering
- Using MapReduce for batch data processing
- Graph based traversals for recommendation systems
- Practical A/B testing and meaningful statistical significance
- Black box learning networks

Session 3. Engineering Solution - 6 hours
In the third session we'll learn enough about implementing solutions at scale that we will at least understand some of the practical challenges working with big data. We will cover various data stores, and tools and frameworks for batch and real time data processing such as Cassandra, Hadoop, Kafka and Storm. We will cover:

- Engineering approaches to big data – using Cassandra and Hadoop to store and process big data
- Using graph databases to get more business value from your data
- Infrastructure for big data – cloud computing and Infrastructure-as-a-Service using tools like Chef
- Real time data processing – the challenge of near real time analytics and engineering solutions using Kafka and Storm

* The case studies will be drawn from a survey of practicing data scientists in the fall based on problems they are solving right now. As such, the details of the problems, algorithms and engineering solutions are subject to change.

ASSIGNMENTS/METHOD OF EVALUATION

Your grade for this course will be based on a combination of quiz’s, individual case studies and group work.

The online quiz’s will be designed to test a combination of retention and understanding of the underlying concepts. The individual case studies will be graded 75% based on the quality of the recommendations and 25% on the clarity of the
written communication. The group projects will be graded equally based upon the quality of the output/outcomes and the use of the concepts from the class to collaborate effectively.

REQUIRED COURSE MATERIALS

None. I’ll share resources through Canvas.
CLASSROOM NORMS AND EXPECTATIONS

Because of the small number of sessions, attendance at all three sessions is required. Eating/drinking in class is allowed and laptop use is encouraged. Class will start on time, so please try to be in the classroom 10 minutes before the scheduled start time.

INCLUSION, ACCOMMODATIONS, AND SUPPORT FOR STUDENTS

At Columbia Business School, we believe that diversity strengthens any community or business model and brings it greater success. Columbia Business School is committed to providing all students with the equal opportunity to thrive in the classroom by providing a learning, living, and working environment free from discrimination, harassment, and bias on the basis of gender, sexual orientation, race, ethnicity, socioeconomic status, or ability.

Students seeking accommodation in the classroom may obtain information on the services offered by Columbia University’s Office of Disability Services online at www.health.columbia.edu/docs/services/ods/index.html or by contacting (212) 854-2388.