

CSCI 490/680 – Big Data Analytics

“Every day, we create 2.5 quintillion bytes of data—so much that 90% of the data in the world today has been created in the last two years alone.”

– IBM¹

The buildup of global data is happening at orders of magnitude that border on the incomprehensible, and the task of mining that data for useful information is becoming exponentially more difficult. To help meet this challenge, the White House launched its “Big Data Research and Development Initiative” in 2012, and Google, Facebook, LinkedIn, Amazon, and other organizations with a major presence in the online space have invested in large-scale data analytics to extract information from massive datasets. Given this environment, the data scientist area has become one of America’s most robust job sectors (see Figure 1).²



Figure 1: Growth in data science jobs

Today, supply is falling far short of satisfying incremental demand for data scientists. As the McKinsey Global Institute reports, “By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions.”³

With this resource gap in mind, this course surveys the foundations of data science and big data, reviews relevant research, and introduces the algorithms and methods used to derive valuable predictions and insights from data. Merging theory with practice, the class covers foundational topics while providing hands-on practical experience with useful languages, toolkits, and frameworks (e.g., Hadoop, Mahout and Spark). The course topics include, but are not limited to: big data management and processing techniques, data analytics, social media mining, recommendation systems, statistical methods, and models. Welcome aboard!

¹ <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>

² <http://money.cnn.com/gallery/pf/2012/11/01/best-new-jobs-in-america/3.html>

³ http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation