Data Science 601

In recent years, data science programs have been proliferating. On September 2015, the University of Michigan announced a $100 million Data Science Initiative (DSI). In their definition of DSI they use words, such as “processing”, “analysis”, and “interpretation of vast amount of data”.

Many statisticians are puzzled by this new discipline which seems to claim to do the same tasks that have been part of their daily work for decades. Also, as large as the UM initiative was statisticians had an insignificant presence which left them marginalized and confused.

When searching the web definition of data science and statistics; while different words are used they clearly overlap. Also, to the argument of data, statisticians have been using large amounts of data of all types for decades. So, for statisticians their profession is just being nicely package as the new and shining Data Science. Many statistics organizations are asking, among other questions: Aren’t we Data Science? Others are raising questions on the relevancy of statistics in relation to data science by making statements such as: Data Science without Statistics is possible even desirable.

For sure, there are many points of views contrasting Data Science and Statistics. Many memes are being repeated. The first one is the need to handle ‘Big Data’. This claim is easily rebutted as statisticians have been handling large datasets very efficiently. The second one is the ‘skill meme’, stating that traditional resources could not accommodate big data; thus, the need of new skills. The third one is the ‘job meme’, which was born with the existence of large global IT organizations; which created a large demand for database and statistics skills.

Beside the excitement there is a lot of confusion surrounding the new data science program. One consensus is that there is a real need for a workforce that can explore data, apply statistical analysis techniques, and write computer programs; as well as understand the business aspect of an organization. The challenge is for universities to satisfy the demand by providing programs that train students in data analysis, statistics, and information technology.