Exploring and Understanding 'Big Data' Through Case Studies

Jerry Platt
Akita International University, Japan

The buzz term for commerce in 2012 is 'Big Data'. As one example, Facebook processes 2.5 billion items, including more than 300 million photos, and more than 500 terabytes of data *every day*. More than sheer volume, 'Big Data' is characterized by its velocity and confounding variety, resulting in discrepancies in value as firms struggle to turn this digital landscape into actionable and profitable information.

Typical business statistics courses, even at the more advanced levels, still rely on toy datasets that, while they may facilitate comprehension of basic concepts, leave graduates woefully unprepared for the volume, velocity, and variety they are likely to encounter in the corporate world. This paper considers a data-driven business curriculum that exposes students to both real-world 'Big Data' and tools to explore and uncover hidden structures and relationships. In particular, a freely-available open-source data mining environment is used to analyze a variety of complete 'Big Data' sets that real corporations have made available to the public in hopes that prize money they offer will entice outsiders to improve on their own internal analyses of the data. Such a challenge stimulates the competitive nature of students.

Several examples of current competitions and sample analyses are demonstrated.