Module IV

Data Access and Data Protection

INFO 590: Applied Data Science
Summer 2017

Data Access

Most of the times the data required for analysis is not readily available at a single source. Even if it is available, often times it is not available in the desired format that makes processing easier. Thus, the data collected from various sources have to aggregated and fit into an appropriate format before it can be used. The video tutorial available in Canvas walks you through the steps followed to create the 'Cricket Dataset' used in Assignment 3.

Data Protection

With the increase in data being generated it becomes essential to have a working system in place to regulate and monitor the access of data. Historic data proves to be an asset to industries in understanding past behaviour and forecasting future demands. Data is considered an asset and the data user is required to abide by applicable laws, policies, procedures and guidelines with respect to access, use, or disclosure of information. While performing data analysis, one often deals with customer data which is considered sensitive information. It is important to respect the confidentiality and privacy of individuals whose records they access.

Tutorials

- **HRMS**

  The Human Resource Management System (HRMS) contains a variety of information about faculty and staff at Indiana University related to their jobs, personal information (such as address and phone numbers), and benefit related enrolment information. This tutorial elaborates on the classification of data and the guidelines data handling/ sharing and data disposal.

- **FERPA**

  Family Educational Rights and Privacy Act (FERPA) protects information about the student which will not be shared with the faculty without the consent of the student. It not just includes critical information like Social Security number or Driver's license number but also includes details like phone, race/ethnicity, gender, country of citizenship, religious affiliation, disciplinary status, marital status etc.

Tutorials

- **Acceptable Use Agreement**

  This agreement outlines the policies and regulations for accessing institutional data and Information technology resources. Users of the Indiana University information and information technology resources must accept certain responsibilities and standards for computer use to be entrusted with computer accounts.

- **Ethical Issues in Data Science**

  This video tutorial is a webinar presented by McGill University Centre of Genomics & Policy Director Dr. Bartha Maria Knoppers. It is part of the web series BD2K Guide to the Fundamentals of Data Science. In this tutorial, she talks about the various approaches namely: Current /self-regulatory approach, Commons approach, Future Governance approach and the Human Rights approach.