Course Information

Description: This course is designed to provide participants with a detailed description of data analytics in health care. Big data in health care has sparked a lot of interest in applying large-scale data analytical techniques for acquiring new information about existing data. In order to become proficient in data analytics one must first understand the foundational component of how the data is stored and acquired. Therefore, the first part of the course will focus on database design and executing structured query language (SQL) scripts in MySQL workbench for acquiring data. Once the data is in hand, the next step is to analyze that data. Commonly, statistical techniques are utilized for testing hypothesis and determining if there are statistically significant observations. However, with the expansion of big data there is an opportunity to move away from hypothesis testing and acquire new knowledge using data mining modeling. In this course, statistical analyses and data mining techniques will be discussed along with methods for deploying these techniques using the open-access analytical software, R. After taking this course, you will have a better understanding of the nature of big data and the methods used for acquiring, analyzing, and ultimately discovering new information from data.

Time Commitment: 60 hours
Pre-Requisites: Basic understanding of health care
Day and Time: Online
Location: Online

Instructor Information

Instructor: David Marc, MBS
Phone: 715-417-0859
E-mail: dmarc@css.edu
Web page: 
Office Hours: Students may communicate with faculty via Bulletin Board or e-mail as frequently as desired by the individual student. Faculty will generally respond within 24-48 hours of receipt of a communication.

Required Materials

Handout Materials: Included in the online unit summaries

Course Outcomes

Upon completion of this course, a participant will be able to:
1. Describe the opportunities and challenges with big data
2. Design a database and execute SQL scripts for querying the data
3. Apply basic statistical and data mining procedures to health care data.
4. Utilize the program R for statistical and data mining purposes.

Course Outline

I. Big Data!
II. An introduction to databases
III. MySQL Workbench
IV. Statistical analysis
V. Data mining
VI. R!
## Assessment Activities

<table>
<thead>
<tr>
<th>Unit</th>
<th>Topics</th>
<th>Outcome(s) Addressed</th>
</tr>
</thead>
</table>
| 1    | Big Data!  
• Read, “The Promise and Peril of Big Data”, by Bollier  
• Unit 1 Lecture | 1 |
| 2    | An Introduction to databases  
• Unit 2 Lecture | 2 |
| 3    | MySQL Workbench  
• Unit 3 Lecture  
• Assessment 1: SQL Assignment | 2 |
| 4    | Statistical Analysis  
• Unit 4 Lecture  
• Assessment 2: Statistics Assignment | 3 |
| 5    | Data Mining  
• Read, “Data Mining Applications in Healthcare”, by Koh & Tan  
• Read, “Application of Data Mining Techniques to Healthcare Data”, by Obenshain  
• Read, “Empirical Study on Applications of Data Mining Techniques in Healthcare”, by Kaur & Wasan  
• Unit 5 Lecture  
• Assessment 3: Data mining Assignment | 3 |
| 6    | R!  
• Unit 6 Lecture  
• Assessment 4: R Assignment | 4 |

### Assignments:

#### MySQL Assignment  
(10 points)  
Choose the best multiple-choice answer for the SQL query that is most appropriate for acquiring the data listed in each problem. 10 questions, 1 point each

#### Statistics Assignment  
(10 points)  
Choose the best multiple-choice answer for the statistical procedure that should be utilized to analyze the data that is acquired in each of the scenarios described. 10 questions, 1 point each

#### Data mining Assignment  
(10 points)  
Choose the best multiple-choice answer for questions related to data preparation, modeling, and evaluation when conducting data mining. 10 questions, 1 point each

#### R Assignment  
(20 points)  
There are two parts to this assignment. In part A, you will be provided a scenario where you must choose the correct R scripts from multiple-choice answers for carrying out a statistical procedure. In part B, the scenario expands to carrying out a data-mining task in R and you must choose the most appropriate multiple-choice answer for each question.