CS 491-4/591-2 Big Data Syllabus
Fall 2014

Instructor: Dr. Mingxuan Sun  
Office: FANER 2138
Email: msun@siu.edu  
Office Phone: 453-6054

Office Hours: WF, 2:00 – 5:00 pm or by appointment
Lecture: MWF, 5:00 – 5:50 pm, FANER 1326
Lab: N/A

Course Objectives: To understand the principles of data analytics and gain the skills to implement a pipeline of data collection, data analysis and data visualization at scale.


Topics:
- Introduction to data analytics
- Classification (k-nearest neighbor, decision tree, logistic regression)
- Linear regression
- Clustering (k means, hierarchical clustering)
- Association rule mining
- Dimension reduction (techniques, visualization)
- Scaling up (Hadoop, Hive)
- Recommendation systems
- Text mining
- Graph analytics

Grading/Evaluation:

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>15%</td>
</tr>
<tr>
<td>Project</td>
<td>25%</td>
</tr>
<tr>
<td>Midterm</td>
<td>30%</td>
</tr>
<tr>
<td>Final</td>
<td>30%, Monday Dec. 8, 6:00 – 7:45 pm</td>
</tr>
</tbody>
</table>

A: 90-100; B: 80-89; C: 70-79; D: 60-69; F: <60

Classroom Policies:

A. Attendance Policy: Students are expected to attend classes regularly. Students are responsible for all announcements made in class and/or posted to D2L.

B. Late Homework/Missed Exams: Late homework is not accepted. You should inform me before the exam if you cannot make it. If an exam is missed for a legitimate reason, a grade will be assigned based on the remaining homework/exams.