

## Subject Description Form

<b>Subject Code</b>	COMP6703
<b>Subject Title</b>	Advanced Topics in Data Analytics
<b>Credit Value</b>	3
<b>Level</b>	6
<b>Pre-requisite / Co-requisite/ Exclusion</b>	Undergraduate introductory background in <ul style="list-style-type: none"> <li>- Probability and statistics</li> <li>- Database and data modeling</li> </ul>
<b>Objectives</b>	<p>The goal of this course is to introduce students to a variety of data analysis methods that are useful for understanding, visualizing and getting insight of data from different researches and applications.</p> <p>In addition to concentrate on formulas and how they are computed, we'll use existing software or write programs to explore a variety of statistical problems concerning text and/or numbers, both numerically and graphically.</p>
<b>Intended Learning Outcomes</b>	<p>Upon completion of the subject, students will be able to:</p> <ul style="list-style-type: none"> <li>(a) understand various statistical methods for data analysis and relate or apply them to the data encountered in research;</li> <li>(b) understand various quantitative methods for data analysis and relate or apply them to the data encountered in research;</li> <li>(c) carry out in-depth analysis of the data encountered in research.</li> </ul>
<b>Subject Synopsis/ Indicative Syllabus</b>	<ol style="list-style-type: none"> <li>1. Data Types and Characteristics               <ol style="list-style-type: none"> <li>a. Relational data, graph data, time series data, text data, survey data, multimedia data, etc.</li> </ol> </li> <li>2. Statistical Methods for Data Analysis               <ol style="list-style-type: none"> <li>a. Multiple, logistic and non-linear regressions</li> <li>b. Discriminant analysis</li> </ol> </li> <li>3. Quantitative Methods for Data Analysis               <ol style="list-style-type: none"> <li>a. Time series analysis</li> <li>b. Probabilistic modeling</li> <li>c. Optimization</li> </ol> </li> <li>4. Decision Analysis               <ol style="list-style-type: none"> <li>a. Multiple objectives</li> <li>b. Decision trees</li> <li>c. Influence</li> <li>d. Sensitivity analysis</li> </ol> </li> <li>5. Exploratory Analysis               <ol style="list-style-type: none"> <li>a. Data Visualization</li> </ol> </li> <li>6. Big Data Analytics               <ol style="list-style-type: none"> <li>a. Unstructured data concepts (key-value)</li> <li>b. MapReduce technology</li> <li>c. Analytics for big data</li> </ol> </li> <li>7. Application Examples</li> </ol>
<b>Teaching/Learning Methodology</b>	Lectures teach students on the main concepts and methods of the course, together with comprehensive examples, and class questions/answers/discussions for easy understanding.

	<p>Tutorials and lab sessions offer the opportunity for students to review and consolidate the lecture and reference materials through exercises and also software tools.</p> <p>Project assignments will give students the opportunity to solve practical data analysis problems.</p> <p>Written assignments help students to develop a solid foundation of data analytics.</p>																																						
<b>Assessment Methods in Alignment with Intended Learning Outcomes</b>	<table border="1"> <thead> <tr> <th rowspan="2">Specific assessment methods/tasks</th> <th rowspan="2">% weighting</th> <th colspan="6">Intended subject learning outcomes to be assessed (Please tick as appropriate)</th> </tr> <tr> <th>a</th> <th>b</th> <th>c</th> <th></th> <th></th> <th></th> </tr> </thead> <tbody> <tr> <td>1. Assignments</td> <td rowspan="2">100%</td> <td>√</td> <td>√</td> <td>√</td> <td></td> <td></td> <td></td> </tr> <tr> <td>2. Project</td> <td>√</td> <td>√</td> <td>√</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Total</td> <td>100%</td> <td colspan="6"></td> </tr> </tbody> </table> <p><b>Assignment(s):</b> assessment of the theoretic studies with respect to the understanding of the relevant subject matters including new concepts, algorithms and techniques by proving answers to the assignment questions</p> <p><b>Project:</b> assessment of the ability for problem solving through real case studies and implementation of a prototype system for demonstration, as well as the oral presentation and the report writing</p>		Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)						a	b	c				1. Assignments	100%	√	√	√				2. Project	√	√	√				Total	100%						
	Specific assessment methods/tasks	% weighting			Intended subject learning outcomes to be assessed (Please tick as appropriate)																																		
a			b	c																																			
1. Assignments	100%	√	√	√																																			
2. Project		√	√	√																																			
Total	100%																																						
<b>Student Study Effort Expected</b>	Class contact:																																						
	Lecture/Tutorial/Lab	39 Hrs.																																					
	Other student study effort:																																						
	Reading	40 Hrs.																																					
	Preparing written and project and assignments,	43 Hrs.																																					
	Total student study effort	122 Hrs.																																					
<b>Reading List and References</b>	<ol style="list-style-type: none"> <li>1. Wikibook: Statistics - Probability and Data Analysis (<a href="http://en.wikibooks.org/wiki/Statistics">http://en.wikibooks.org/wiki/Statistics</a>)</li> <li>2. John A. Rice, Mathematical Statistics and Data Analysis (with CD Data Sets) (Duxbury Advanced), 3<sup>rd</sup> Ed., 2006.</li> <li>3. Philipp K. Janert, Data Analysis with Open Source Tools (A hands-on guide for programmers and data scientists), O'Reilly Media, 2010.</li> <li>4. Jimmy Lin and Chris Dyer, Data-Intensive Text Processing with MapReduce, Morgan &amp; Claypool Publishers, 2010.</li> </ol> <p>+ web references</p>																																						