



## INF410: Data Analytics for Business Summer 2018

### Course Details

Course Title	Data Analytics for Business
Course Code	INF410
Credit Hours	3
Semester & Year	Summer 2018
Pre/Co-requisites	Basic Statistical Methods, Quantitative Methods
Department	Management Science
Course Coordinator	Ms. Lee Ji Won


### Contact Hours

Class Type	Days	Time	Room
Lecture	Mon-Fri	9 am – 12pm	601

### Instructor Details

Instructor	Tahir Hameed
Room	1105, SolBridge
Consultation Hours	By appointment
Email	tahir@solbridge.ac.kr
Telephone	630-8562
Course Webpage	<a href="http://smart.wsu.ac.kr">http://smart.wsu.ac.kr</a>

### Mission Map

	Mission Based Goals	Approximate % of Course Content	Approximate % of Assessment
	Global Perspective	50%	50%
	Asian Expertise	10%	10%
	Creative Management Mind	30%	30%
	Cross Cultural Competence	0%	0%
	Social Responsibility	10%	10%
	<b>Total</b>	<b>100%</b>	<b>100%</b>

## Course Objectives

Digitalization has transformed business management in big ways. These days most business transactions, operations and communications are being conducted online. Employees, suppliers and customers are interacting with each other in unprecedented ways and scales. As a result, new generation enterprise systems, social networking sites, cloud and mobile-based apps are generating troves of business data (BIG DATA). Firms are still grappling with the challenges related to these newer, bigger and fast-changing datasets. The demand for data modelers, business analysts and data scientists – capable of mining actionable intelligence from big data – has outpaced the number of available professionals.

This course aims to introduce participants to data science and applications of big data analytics technologies for business decision support. It would provide a helicopter view of the full data analytical cycle to the participants involving problem definition, data mining, modeling, data analysis and its application to business decision-making. Along the way, it will also equip the participants with most relevant tools and skills in each of the phases of the data analytic cycle. Although the course does not involve or require programming knowledge neither does it goes deep into technologies like in-memory databases, artificial intelligence and machine learning, owing to the technical nature of the subject, participants are expected to be willing to invest time in learning several supporting concepts and tools to make full sense of the discussed topics.

To help the participants get the gist of applications of data analytics to business, at least three hands-on cases/exercises will be completed by each participant using real datasets from retail, marketing, finance and/or healthcare sectors. For instance, profiling of customers and their spending behaviors helps retailers in launching targeted marketing and advertising strategies for specific customer segments. In the healthcare sector, prediction of high-risk patients helps healthcare organizations to allocate their limited resources and interventions to those patients who could benefit the most. Credit card firms, insurance companies and banks commonly use context analysis for fraud detection to prevent or minimize their financial losses from fraudulent transactions.

## Learning Goals

Upon completion of the course participants should be able to:

1. Demonstrate good understanding of how big data technologies impact business processes and key management functions
2. Demonstrate basic knowledge of full data analytics lifecycle and key tasks in each phase
3. Recognize situations where different modelling and business analytics techniques could contribute value in business decision-making
4. Evaluate the suitability of statistical methods for types of business decisions, for example decision trees, logistic regression, and time series analysis
5. Evaluate and apply suitable data mining, AI and machine learning techniques to specific business problems/domains, for example
  - a. Sentiment analysis and text mining to customer reviews for marketing decisions
  - b. Social network analysis for organizational change
  - c. Context analysis of financial transactions for risk and/or fraud detection
6. Plan and execute a small business data analysis project from problem definition to testing and operationalization of the proposed analytical model

## Teaching Methodology

This course will be taught through a series of lectures, in-class (group and individual) exercises and take-home assignments. All teaching and reference materials will be posted on SMART LMS. Participants will also submit their assignments online. However, hard copies for most submissions will also be required on announced due dates. Students are advised to check their SolBridge email and SMART LMS system frequently.

## Ethics

### Plagiarism

Plagiarism will not be tolerated in any form. Penalties can be as severe as expulsion from the university. To avoid plagiarism, it is always best to do your own work or cite the work of others appropriately. Refer to your student handbook for detailed description of plagiarism and penalties.

For this class:

1. The first instance of plagiarism will result in a “zero” for the assignment in question, and a report will be filed with disciplinary officer.
2. The second instance of plagiarism will result in a fail grade for the entire course, and a report will be filed with disciplinary officer.
3. The third instance will result in institutional-level disciplinary action which could include expulsion from the school.
4. The instructor will report each instance of plagiarism, academic dishonesty and violation of school disciplinary rules to the disciplinary officer.

### Copying Textbooks, copyrighted materials and academic dishonesty

Copying textbooks and other copyrighted materials without permission of publisher or author is tantamount to theft. Therefore, students are expected to purchase the prescribed books and other materials from university bookstore or elsewhere. The institutional policy is that

- Students using copied versions of books without permission will be asked to leave the class.
- They may get “zero” participation points.

**Academic Dishonesty** includes but not limited to: (a) cheating during examinations, (b) obtaining/providing information for reports, assignments and examinations by fraudulent and unauthorized means, and (c) false representation of others’ effort as one’s own. Please refer to your student handbook for details and penalties.

## Course Materials, Software and Resources

Selected chapters and exercises will be assigned from main and/or secondary textbooks.

### Main textbook

Title	Data Science for Business
Edition	2nd Edition
Authors	Foster Provost and Tom Fawcett
Publisher	O’Reilly Media
ISBN 13	978-1449361327

### Secondary (Reference) Textbook

Title	Essentials of Business Analytics
Edition	2017
Authors	Jeffrey D. Camm, James J. Cochran, Michael J. Fry, Jeffrey W. Ohlmann, David R. Anderson, Dennis J. Sweeney, Thomas A. Williams
Publisher	Cengage
ISBN 13	978-1285187273

### **Supplementary texts**

1. A complete online text for statistical analysis concepts: <https://www.statsoft.com/Textbook>

### **MOOCs/Video on Demand:**

#### **Required:**

2. Babson Course on edX.org - Analytics for Decision Making :  
<https://courses.edx.org/courses/course-v1:BabsonX+BPET.STATx+2T2017/course/>
3. Columbia Course on edX.org - Statistical Thinking for Data Science and Analytics  
<https://courses.edx.org/courses/course-v1:ColumbiaX+DS101X+1T2016/course/>
4. Orange Canvas Channel: <https://www.youtube.com/channel/UCIKKWBe2SCAEyv7ZNGhle4g>
5. Microsoft Power BI: <https://www.youtube.com/user/mspowerbi> and  
<https://www.youtube.com/watch?v=BsXliHbOFDM>

#### **Others:**

6. Recap Statistics and Probability: <https://www.khanacademy.org/math/statistics-probability>
7. Linear Regression: <https://www.youtube.com/watch?v=ZkjP5RJLQF4>
8. Multi-logistic regression: <https://www.youtube.com/watch?v=zAULhNrnuL4>
9. Time-Series Analysis: [https://www.youtube.com/watch?v=Prpu\\_U5tKkE&t=1188s](https://www.youtube.com/watch?v=Prpu_U5tKkE&t=1188s) and  
<https://www.youtube.com/watch?v=HIWXdHIDSFs>
10. Decision Tree Classification for predictions:  
<https://www.youtube.com/watch?v=jzoDKtnTPpg> and  
<https://www.youtube.com/watch?v=eKD5gxPPeY0>
11. MOOC channel on WEKA Machine Learning: <https://www.youtube.com/user/WekaMOOC>
12. ESRI MOOC The Location Advantage <http://www.esri.com/mooc/location-advantage>

### **Software and Tools**

#### **Required:**

13. Orange Canvas data mining & visualization software, download: <https://orange.biolab.si/>
14. Microsoft Power BI for Business Intelligence: <https://powerbi.microsoft.com/en-us/>
15. Weka Machine Learning: <https://www.cs.waikato.ac.nz/ml/weka/>
16. Microsoft Access or MySQL (depending on availability and/or the datasets included)
17. Microsoft Excel with some add-ons (to be announced in class)

#### **Others:**

18. NodeXL; social network analysis add-on for Microsoft Excel
19. AutoMap text mining – download: <http://www.casos.cs.cmu.edu/projects/automap/>
20. ORA-Lite network visualization - download: <http://www.casos.cs.cmu.edu/projects/ora/>

### **Datasets and Cases:**

#### **Required:**

21. Primary textbook. Join Google Group: <https://groups.google.com/forum/#!forum/data-science-for-biz>
22. Reference textbook datasets: Will be distributed in class ([http://www.cengage.com/cgi-wadsworth/course\\_products\\_wp.pl?fid=M20bl&product\\_isbn\\_issn=9781305627734](http://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20bl&product_isbn_issn=9781305627734))

#### **For Reference:**

23. UCI Machine Learning: <https://archive.ics.uci.edu/ml/datasets.html>
24. Stanford Large Network Dataset Collection (SNAP): <https://snap.stanford.edu/data/>
25. Arizona State University <http://socialcomputing.asu.edu/pages/datasets>
26. Healthcare MIMIC2 dataset: <https://physionet.org/mimic2/>
27. Chicago City Data Portal: <https://data.cityofchicago.org/>
28. UK Maps datasets by Ordnance Survey: <https://www.ordnancesurvey.co.uk/business-and->

[government/products/opendata-products.html](http://government/products/opendata-products.html)

29. Global open data sets: <https://index.okfn.org/dataset/postcodes/>

30. Edinburgh: <http://www.inf.ed.ac.uk/teaching/courses/dme/html/datasets0405.html>

31. BDA competitions/challenges: <https://www.kaggle.com/competitions>

## Assessment

There are several assessment venues in this class.

Component	Actual Weight	For Grading
1. Class Attendance (school policy)	20%	Attendance 20%
2. In-class or take-home exercises (statistical methods, data mining, data analysis techniques)	20%	Assignments 20%
3. Mid-Term Exam – Data Visualization Project	20%	Mid-Term Exam 20%
4. Final Exam-Part 1 – Data Analysis Project/s	30%	Final Exam 40%
5. Final Exam-Part 2 – Class Participation	10%	
<b>Total</b>	<b>100%</b>	<b>100%</b>

### Class Attendance (20%)

Class attendance bears 20%. Pl. note that 3 hours of non-attendance could cost you a letter grade change.

### Assignments (Statistics, Data Mining, and Modelling Prep Exercises): (20%)

In-class or take-home exercise requires each student to learn to apply statistical methods or tools for data mining and preparation, loading, modelling and analysis. The foundation knowledge and skills will prepare the students to take up integrated projects and exercises.

### Midterm Examination (Data Visualization Project) (20%)

The students will be provided an Excel or (other format) database. A managerial dashboard for decision support purposes should be completed as part of the data visualization project. Microsoft Power BI is the preferred software for development of the dashboard but in the interest of time and capabilities, Microsoft Excel might also be used after discussion with instructor.

### Final Examination-Part 1: Individual Project/s or Case Study Analyses (30%)

There will be two parts of the Final Exam. Part 1 will comprise of one large or three small business analytics projects/case studies. A data analytics project entails understanding and implementation of full data analysis cycle i.e. problem definition (business case preparation), data mining, data modelling, data analysis and decision-making. Each phase will be described in writing as part of a report or written analysis of a case. If the course is offered in a short semester, case studies with already mined datasets and hints for modeling could be used by the instructor to speed up progress. Written reports would be submitted and/or presented. Detailed requirements would be provided in the class.

### Final Examination Part 2 Class Participation (10%)

Class Participation would mainly be reflections about the sessions and any contributions in or out of the class. All of your contributions will be noticed and awarded.

## Course Schedule

Week-Lec-Date	Topics & Textbook Readings	Assignments and Practice
<b>Part I: Introduction and background knowledge</b>		
<b>Week I – Lec 1 and 2</b>	<b>Introduction:</b> Big data analytics, data-analytical thinking and business analytics. syllabus	Articles, Videos, Exercises
<b>Week II – Lec 3 and 4</b>	<b>Recap: Statistics and Probability</b> - statistics, estimation, inference, hypothesis testing	Texts, Video Lectures, In-class and Take-Home Exercise
<b>Week III – Lec 5 and 6</b>	<b>Recap Algos</b> – Linear Regression, Decision Trees, Logistic regression, Clustering, Time Series	Texts, Video Lectures, In-class and Take-Home Exercise
<b>Week IV – Lec 7 and 8</b>	<b>Getting Ready:</b> Installing Orange Canvas, RStudio, Microsoft BI and Weka software	Texts, Video Lectures, In-class and Take-Home Exercise
<b>Week V – Lec 9 and 10</b>	<b>Exploratory Data Analysis and Data Visualization</b> - Mid-term project launch	Texts, Video Lectures, In-class and Take-Home Exercise
<b>Part II : Data Analytics Cycle</b>		
<b>Week VI – Lec 11 and 12</b>	<b>Data Analytics Life Cycle.</b> Introducing the data analytics case analysis template	Texts, Video Lectures, In-class and Take-Home Exercise
<b>Week VII – Lec 13 and 14</b>	<b>Data Mining and Predictive Modelling</b> – Unsupervised Learning: K-means clustering, Time Series Analysis	Texts, Video Lectures, In-class and Take-Home Exercise
<b>Week VIII – Lec 15 and 16</b>	<b>Data Mining and Predictive Modelling</b> – Supervised Learning: Decision Trees, Multi-Logistic Regression	Texts, Video Lectures, In-class and Take-Home Exercise
<b>Week IX – Lec 17 and 18</b>	<b>Model Evaluation and Performance</b> –	Texts, Video Lectures, In-class and Take-Home Exercise
<b>Week X – Lec 19 and 20</b>	<b>Mid-Term Exam Due</b>	
<b>Part III: Business Analytics</b>		
<b>Week XI – Lec 21 and 22</b>	<b>Case Study 1:</b> Sentiment Analysis and applications to Marketing	Video Lectures, In-class and Take-Home Exercise
<b>Week XII – Lec 23 and 24</b>	<b>Case Study 2:</b> Context analysis (time-series analysis) for fraud detection in financial transactions	Video Lectures, In-class and Take-Home Exercise
<b>Week XIII – Lec 25 and 26</b>	<b>Case Study 3:</b> Social Network Analysis for people and organizational change	Video Lectures, In-class and Take-Home Exercise
<b>Week XIV – Lec 27 and 28</b>	<b>Case Study 4 (Depending on Time Availability):</b> Predicting 30-days Patient Readmission and optimizing healthcare resources	
<b>Week XV – Lec 29 and 30</b>	<b>Wrap up</b> and/or Presentations -	Submissions of all case study analyses/projects