UCD Michael Smurfit
Graduate Business School

MKT 46090: Big Data Analytics

Fall Trimester 2023/24

Time (Room): Wednesday 15:00-18:00 (N203-GSB)

Office Hours (Room): By appointment

Instructor: David DeFranza
Email: david.defranza@ucd.ie

Every day, an increasing variety of new information is created at a larger volume and faster velocity than ever before. This information presents incredible opportunities for businesses but contending with its size and speed poses considerable technical and strategic challenges. Indeed, businesses capable of extracting knowledge from large data sets can achieve a considerable competitive advantage, and managers capable of facilitating this process will find they have an advantage in the job market. In this module, we will engage with business problems using data analytic thinking. Along the way, we will discuss the fundamental principles guiding the extraction of knowledge from information, challenges posed by very large data sets (including "Big Data"), and some of the most common techniques and technologies used to mine such data.

Learning Outcomes

This module aims to prepare you for the challenges and opportunities today's business leaders face when dealing with "Big Data" and extracting knowledge from large data sets more generally. At the end of this module, you should be able to:

1. Explain the unique characteristics of and challenges posed by Big Data
2. Summarize the industry standard data mining process within the context of a business problem
3. Identify an appropriate analysis method based on a description of the business problem and available data
4. Assess the performance of a model or analysis based on common diagnostic metrics
5. Develop a comprehensive data analysis proposal which addresses a specific business problem

Required Readings

This course will utilize two textbooks, readings from which are required. The first book is available from the library and new, used, and electronic editions are available from most book sellers. The second book is available as an electronic (PDF) edition. The books are:

Provost, F., & Fawcett, T. (2013) *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly.

- Publisher link: https://www.oreilly.com/library/view/data-science-for/9781449374273/
- UCD library link: https://tinyurl.com/b873bppp

Zhou, H. (2020). *Learn data mining through Excel: A step-by-step approach for understanding machine learning methods*. Springer.

- Publisher link: https://link.springer.com/book/10.1007/978-1-4842-5982-5

Assigned chapters as well as required and recommended articles for each week will be posted to Brightspace.

Cases are provided through Harvard Business Publishing. Please use the links below to download a copy of the cases. Note that many cases include a supplemental data file which you should also download.

Title	Authors	URL
Tapping into a Digital Brain: AI-Powered Talent Management at Infosys	Mohanbir Sawhney, Varun Poddar, Charlotte Snyder	https://hbsp.harvard.edu/tu/c63a8a4e
Predicting Consumer Tastes with Big Data at Gap	Ayelet Israeli, Jill Avery	https://hbsp.harvard.edu/tu/30792b5b
Luminar: Leveraging Big Data Using Corporate Entrepreneurship	Simon Parker, Chandra Sekhar Ramasastry	https://hbsp.harvard.edu/tu/28e0151d
Assessing Prediction Accuracy of Machine Learning Models	Michael W. Toffel, Natalie Epstein, Kris Ferreira, Yael Grushka-Cockayne	https://hbsp.harvard.edu/tu/d46b0c0a
Improving Strategic Execution With Machine Learning	Michael Schrage, David Kiron	https://hbsp.harvard.edu/tu/ce6bb1f6

Required Software

This module assumes access to and competency with spreadsheet software such as Microsoft Excel or Google Sheets. Please note that both tools are available for free to students through UCD Connect. Applications will be discussed specifically in Microsoft Excel, which is recommended.

If you are not comfortable with the basic functions of Microsoft Excel (particularly cell references and functions) you can review these concepts through DataCamp courses which are provided as part of this module. Please note that completion of these courses is optional, and your progress will not be monitored or evaluated by instructors.

You can access Data Camp via:
https://www.datacamp.com/groups/shared_links/5bd3e8767880dfbd8ad65b18e9a0285ac2e94ab0d3b34e9142365f36a401df2e

Assessment

This module has three assessment components, each with a specific weighting and marks totally 100%. There will be no exam. The weighting assigned to and responsibility for (i.e., whether an assessment is to be completed by an individual or group) each component is shown below. Following, each assessment component is discussed in detail.

Assessment Component	Weighting	Individual / Group	Deadline
1. In-Class Quizzes	35%	Individual	25 October & 29 November
2. Case Analysis	20%	Individual	17 October
3. Class Participation	5%	Individual	23:59 after each lecture
4. Data Analysis Proposal	40%	Group	8 December

In-Class Quizzes

In-class quizzes will test your knowledge of the course material. For each quiz you will have to answer between 15 and 20 multiple-choice type questions about the content covered in the preceding weeks. All quizzes will be conducted in-class, during the first 30 minutes of class.

Please note that quizzes will only be available for the time specified and that failure to conduct the quiz will automatically result in zero marks. During that window you a limited amount of time to correctly answer as many questions as possible. Please note that you will answer a small subset taken from a large pool of questions, and that the questions will be randomly allocated to each student. This means that students are unlikely to see the same questions and/or questions in the same order. Quizzes may cover any material discussed in class or in the assigned reading.

Case Analysis

Case studies and articles will be used throughout the course to frame discussion and illustrate key concepts related to Big Data and machine learning management. However, one case will be

assigned for formal analysis. Students will work individually to answer a set of questions based on the information in the case and concepts we have discussed in class up to that point. Written analyses will be submitted via Brightspace no later than 23:59 on 17 October. Case analyses should be formatted with 12-point font, single-spaced, with 2.54 cm margins and not exceed four pages. Additional details will be provided via Brightspace.

Class Participation

Students are, of course, encouraged to participate actively during lectures by asking and/or answering questions. Formal assessment of class participation will take place through the submission of regular "minute essays." After each lecture, an assignment will be made available on Brightspace. This assignment will consist of three questions: What was the most important or useful thing you learned in today's class; What question(s) do you have about the material covered in today's class; How might the material we discussed in today's class apply to your role as a manager or business leader?

Responses will be due before 23:59 that day. You should be able to answer each question in fewer than 200 words written in only a few minutes. Reponses will be reviewed at the beginning of the following lecture.

Data Analysis Proposal

Working in randomly assigned teams, you will be asked to develop a comprehensive data analysis plan addressing a specific business problem. The problem is detailed in the required case "Predicting Consumer Tastes with Big Data at Gap." In this case, you will learn that the new CEO of Gap has decided to upend the creative process at the clothing brand by replacing traditional creative directors with data analysts. As managers, it will be your responsibility to lead the team tasked with predicting fashion trends and developing new products using data and analytics. Your team must craft a plan capable of addressing this considerable challenge. Your analysis plan should follow the template provided by Provost and Fawcett (2013; see above), the CRISP-DM framework, and other strategic frames discussed in class. Final reports should be submitted via Brightspace no later than 1 December, 2023. An opportunity for peer- and self-evaluation will be provided. The assignment will be graded in accordance with a rubric which will be provided in advance. Additional details will be provided in class and via Brightspace.

Assessment Criteria and Grade Descriptors

This module utilizes criterion referencing and UCD grade descriptors. Before attempting the assessments for this module, you are encouraged to review the grade descriptors. A copy of the UCD grade descriptors can be downloaded from:

https://www.ucd.ie/history/t4media/UCD%20Module%20Grade%20Descriptors-1.pdf

Protocol for submitting your assignments: All continuous assessment should be submitted electronically via Brightspace, by the deadline specified. Please do not email assignments directly to the teaching team, unless explicitly directed to do so.

Statement of Inclusion

This module strives to be a model of inclusion. We respect and value student diversity in all of the modules we offer. We aim to provide and promote equitable access and opportunity to all students regardless of disability, race, age, gender, sexuality or socio-economic status. Students are encouraged to approach staff to discuss their learning needs. Any information disclosed will be treated confidentially.

University Policies

You should ensure you are familiar with the following UCD protocols:

- ***Plagiarism and Academic Integrity:*** UCD and the College of Business take academic integrity extremely seriously. All work must be your own, be completed specifically for this module and not have been submitted elsewhere. It should also be accompanied by a signed own work statement, such as the following:

  > *I declare that all materials included in this essay/report/project/dissertation is the end result of my own work and that due acknowledgement have been given in the bibliography and references to ALL sources be they printed, electronic or personal.*

  The university's plagiarism and academic integrity policy is available from: https://www.ucd.ie/secca/studentconduct/

- ***Harvard Referencing Style:*** UCD College of Business uses the Harvard style of referencing. The UCD library has developed some resources on avoiding plagiarism and on how to reference correctly using the Harvard style. These resources are available from: https://libguides.ucd.ie/academicintegrity

- ***Assessment Submission Form:*** When submitting a piece of assessment, you are asked to attach an assessment submission form. This form is available from: https://www.ucd.ie/t4cms/assessment%20submission%20form.pdf

- ***Late Submission of Coursework:*** This policy outlines the steps you should take where you know in advance that you will not be in a position to meet a submission deadline and the penalties imposed in such circumstances. See https://www.ucd.ie/t4cms/latesub_po.pdf

- ***UCD Extenuating Circumstances policy:*** If, during the course of this module, you encounter any serious unforeseen circumstances that are beyond your control and which prevent you from meeting the requirements of the module, you should consult this policy. A student guide to this policy is available from: https://www.ucd.ie/students/studentdesk/extenuatingcircumstances/

- ***UCD Student Code:*** The UCD Student Code establishes the University's regulations and expectations in respect of student behaviour and conduct. The Student Code is available from: https://www.ucd.ie/secca/studentconduct/

Module Topics

The schedule below outlines the planned themes, by lecture. Updates and additions will be notified in class and on Brightspace. Required reading and additional information will be posted to Brightspace.

Lecture	Date	Topic
1	13 September	No class
2	20 September	Managing data science; Applying data analytic thinking
3	27 September	Finding and working with data
4	4 October	Feature importance & feature engineering
5	11 October	Distance- and similarity-based models
	17 October	**Case Analysis Due**
6	18 October	Building predictive models; Intro to tree-based models
7	25 October	**Quiz 1**; Implementing tree-based models I
8	1 November	Implementing tree-based models II
9	8 November	Introduction to linear functions
10	15 November	Logistic regression
11	22 November	Evaluating model performance
12	29 November	**Quiz 2**; Artificial neural networks
	8 December	**Data Analysis Proposal Due**