



UCD Michael Smurfit
Graduate Business School

MKT 46090: Big Data Analytics

Autumn Trimester 2024/25

Time (Room): Wednesday 14:00-16:50 (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. Apply industry standard best practices for the organization and documentation of large datasets
3. Summarize the data mining process within the context of a business problem
4. Identify an appropriate analysis method based on a description of the business problem and available data
5. Assess the performance of a model or analysis based on common diagnostic metrics
6. Explain foundational data mining methods and machine learning algorithms

Required Readings

This course will draw on two textbooks, readings from provide a background for topics discussed in class. 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 eBook from the library or 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://go.exlibris.link/gVGYY4zh>

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>
- UCD library link: <https://go.exlibris.link/WJRJNnGW>

Chapters will be assigned as background reading. **Required reading** for each week will be posted to Brightspace.

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

| Title | Authors |
|--|---|
| Tapping into a Digital Brain: AI-Powered Talent Management at Infosys | Mohanbir Sawhney, Varun Poddar, Charlotte Snyder |
| Predicting Consumer Tastes with Big Data at Gap | Ayelet Israeli, Jill Avery |
| GitHub Copilot: Great Promises Tempered by Looming Ethical Shadows | Parul Gupta |
| Nata Supermarkets: Customer Analytics | Bissan Ghaddar |
| Allianz: Optimizing Customer Acquisition Strategy using Machine Learning | Bjarne Brie, Tineke Distelmans, Kristof Stouthuysen, Tim Verdonck, Christopher Grumiau, Thoppan Sudaman |
| Cluster Analysis for Segmentation | Rajkumar Venkatesan |
| Generative AI Value Chain | Andy Wu, Matt Higgins |
| Assessing Prediction Accuracy of Machine Learning Models | Michael W. Toffel, Natalie Epstein, Kris Ferreira, Yael Grushka-Cockayne |

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 basics 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 DataCamp via the URL provided on Brightspace.

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. Case Analysis | 30% | Individual | 24 September, 8 October, 22 October, 12 November |
| 3. Final Exam | 30% | Individual | TBD (Week 12) |
| 4. Data Analysis Project | 40% | Group | 15 December |

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, four cases 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 the posted due date. Case analyses should be formatted with 12-point font, single-spaced, with 2.54 cm margins and not exceed two pages. Additional details will be provided via Brightspace.

Final Exam

The final exam will be one hour long and consist of up to 50 multiple-choice questions. The exam will be held at the Blackrock Exam Centre (see the end of this document for directions). The exam is closed-book and paper-and-pencil. Additional details will be provided during the term.

Data Analysis Project

Working in randomly assigned teams, you will be asked to conduct a comprehensive data analysis addressing a specific business problem. The problem is detailed in the required case “Allianz: Optimizing Customer Acquisition Strategy using Machine Learning.” In this case, you will learn that online sales for the company have fallen sharply and unexpectedly. Working as marketing analysts, your team must utilize the available data to understand consumer segments, predict future sales, and make recommendations to senior management. Your analysis should follow the template provided by Provost and Fawcett (2013; see above), the CRISP-DM framework, and other strategic frames discussed in class.

One final submission for each group should consist of three artifacts. The first artifact (20% of project grade) is a detailed data sheet and/or dictionary documenting the underlying dataset. The second (40% of project grade) is a well-organized and documented Excel file containing your complete analysis of the data. Finally, the third artifact (40% of project grade) is a presentation (either PDF or PowerPoint format) which carefully details the results of applying the CRISP-DM to the case. The presentation file should be created in the style of a corporate presentation which your team might deliver to senior management.

Fundamentally, the analysis and report should answer the following questions:

1. What is the business problem in this case and how is this problem transformed into a machine learning problem?
2. Who are T&B customers? Describe the most important characteristics of the customers and their outstanding insurance policies.
3. Compare T&B customers to the online visitors that did not convert into purchasing customers. Can you identify the main drivers of customer conversion?
4. Can you identify customer segments using unsupervised learning? Can you predict customer conversion?
5. Evaluate the outcome of the predictive models you used. How effective are they in predicting customer conversion?
6. Develop a customer acquisition strategy for Allianz using the analysis results and insights from assignment questions 2 to 5.

Keep in mind that these questions should be answered within the context of the CRISP-DM.

Final reports and peer-evaluations should be submitted via Brightspace no later than 15 December, 2024. 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/registry/t4media/UCD%20Module%20Grade%20Descriptors.pdf>

Please note that this course utilizes the “**Alternative Linear Conversion Grade Scale**” for conversion of numeric grades to letter grades, including calculation of a final grade from the course average. The scale can be found here:

<https://www.ucd.ie/registry/t4media/Standard%20Conversion%20Grade%20Scale.pdf>

Protocol for submitting your assignments: All continuous assessment should be submitted electronically via Brightspace or handed to the instructor in class, as directed, 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 | 11 September | Course Overview; Intro to Big Data & Machine Learning |
| 2 | 18 September | Finding and Managing Data |
| | 24 September | Case Analysis 1 Due |
| 3 | 25 September | Data Understanding and Quality Assessment |
| 4 | 2 October | Feature Engineering & Applications |
| | 8 October | Case Analysis 2 Due |
| 5 | 9 October | Association Rules & Recommender Systems |
| 6 | 16 October | Clustering Methods & Applications |
| | 22 October | Case Analysis 3 Due |
| 7 | 23 October | Clustering Methods & Applications 2 |
| 8 | 30 October | Introduction to Predictive Modeling |
| 9 | 6 November | Tree-Based Models |
| | 12 November | Case Analysis 4 Due |
| 10 | 13 November | Logistic Regression |
| 11 | 20 November | Support Vector Machines |
| 12 | 27 November | Neural Networks |
| | TBD | Final Exam |
| | 15 December | Data Analysis Project Due |

How to Find D Building Offices



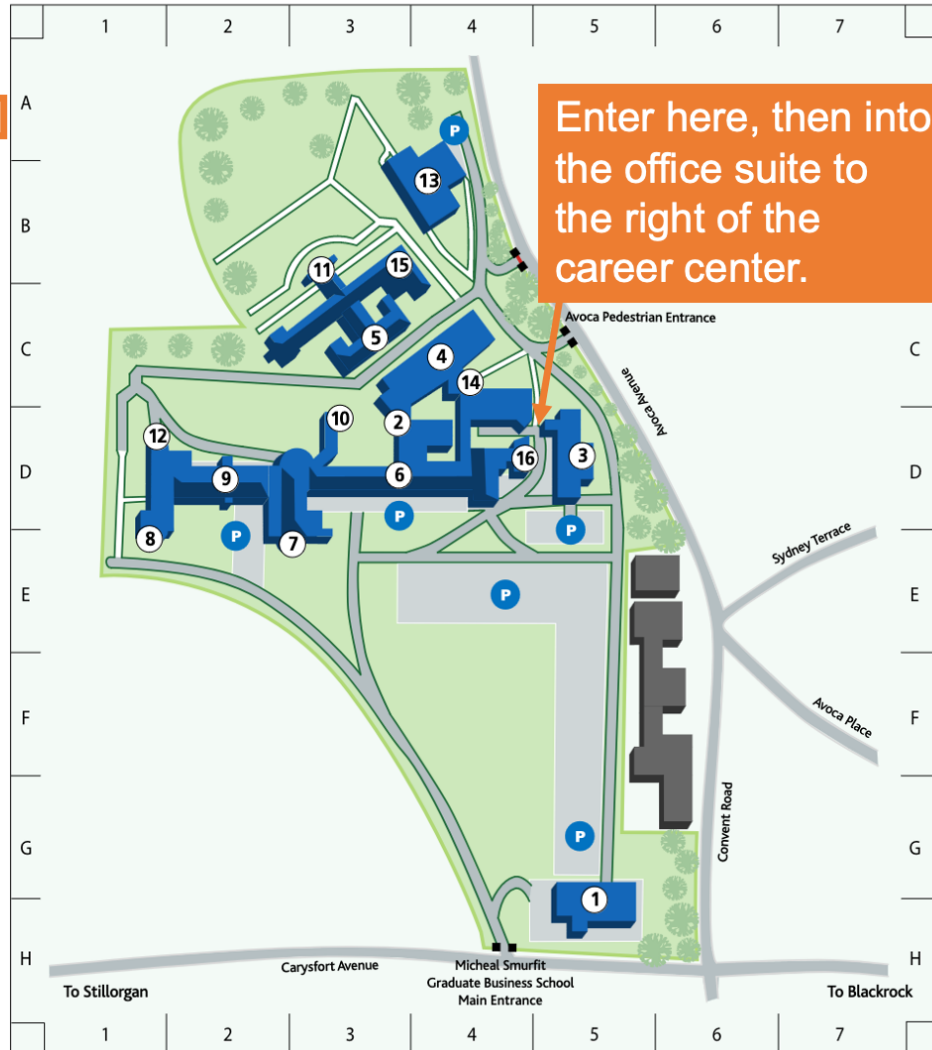
UCD Blackrock Campus

| Building Index | No. | Grid |
|---|-----------|-----------|
| Blackrock Examination Centre | 1. | H5 |
| Copi-Print / ILTG | 2. | D3 |
| D Building | 3. | D5 |
| E Building | 4. | C4 |
| East Hall | 5. | C3 |
| Graduate School of Business | 6. | D3 |
| Library | 7. | E3 |
| Liguori House (Grey House) | 8. | E1 |
| Management House Executive Education | 9. | D2 |
| Marketing Development Programme (Granite Room) | 10. | D3 |
| Oratory | 11. | B3 |
| Proby House | 12. | D1 |
| Restaurant | 13. | B4 |
| Services Desk | 14. | C4 |
| West Hall | 15. | B3 |
| Laundry Room | 16. | D4 |


LEGEND

- Building
- Walkways
- Roadways
- Parking

UCD Unicare:
 Our campus, Our care...
 Emergency Line: (01 716) 7999




How to Find the Blackrock Exam Centre






UCD EXAMS

How to get to the Blackrock Exam Centre

From: UCD Belfield
To: UCD Smurfit School, Blackrock

 **Bus Number S6 then walk for approx 15 min**
From stop no. 765 (Belfield campus) or stop no. 7963 (UCD Village)

Estimated journey time: approx. 35 - 40 minutes (not including waiting time for the bus)

-  Take the number S6 bus from Belfield in the direction of Blackrock and get off the bus in Blackrock Park (stop 3032). Check <https://www.transportforireland.ie/> for timetables.
-  Walk up Carysfort Avenue (see map) and turn right into the UCD Smurfit School. The Blackrock Exam Centre is the building on the right close to Carysfort Avenue.
-  To return to Belfield, take the S6 from Blackrock in the direction of UCD.

MAKE SURE YOU ARRIVE AT LEAST 15 MINS BEFORE YOUR EXAM



Blackrock DART Station → **Blackrock**

Stop no. 3032 Blackrock Park
Get off the bus here and walk to UCD Smurfit.

18 min
1.4 km

14 min
1.1 km

16 min
1.2 km

Carysfort Avenue

UCD Smurfit Exam Hall

For a Google Maps Pin, scan below:

