

2015-4

Lessons Learned from Teaching Data Analytics in a Fully Online Mode at Postgraduate Level

Daniel McSweeney

Institute of Technology, Blanchardstown, daniel.mcsweeney@tudublin.ie

Markus Hoffmann

Institute of Technology, Blanchardstown

Geraldine Gray

Technological University Dublin, geraldine.gray@tudublin.ie

See next page for additional authors

Follow this and additional works at: <https://arrow.tudublin.ie/st3>



Part of the [Higher Education Commons](#)

Recommended Citation

McSweeney, D., Hoffmann, M., Gray, G. & Keyes, L. (2015). Lessons learned from Teaching Data Analytics in a Fully Online Mode at Postgraduate Level. *Higher Education in Transformation Conference, Dublin, Ireland, 2015, pp.265-272.*

This Conference Paper is brought to you for free and open access by the Higher Education in Transformation Conference, Dublin, 2015 at ARROW@TU Dublin. It has been accepted for inclusion in Stream 3: Digital Campus and Universal Design by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, vera.kilshaw@tudublin.ie.

Funder: Tu4D

Authors

Daniel McSweeney, Markus Hoffmann, Geraldine Gray, and Laura Keyes

Lessons Learned From Teaching Data Analytics in a Fully Online Mode at Postgraduate Level

Daniel McSweeney, Markus Hofmann, Geraldine Gray, Laura Keyes

Department of Computing, School of Informatics and Engineering,
Institute of Technology Blanchardstown, Ireland

Abstract

This paper reports on the experiences of developing and delivering the online MSc. in Business Intelligence and Data Mining between 2010 and 2013. We outline our rationale for the selection of Adobe Connect as a synchronous classroom tool and discuss our use of the moodle virtual learning environment to support asynchronous learning and communication. We also discuss some of the key changes in approaches to teaching and assessment and the challenges that were faced by academics in adapting classroom based courses to an online distance education environment. We conclude by examining student interactions with the synchronous and asynchronous delivery technologies and reflect on feedback provided by staff and students throughout the delivery of the course.

Keywords: online distance learning, online assessment, masters in data analytics.

The MSc. in Business Intelligence and Data Mining is a fully online, part-time and applied course, which is accredited by the Institute of Technology Blanchardstown. The programme began in 2010 and was designed in collaboration with industry partners. Graduates are equipped for employment in sectors where data analytics is an essential component. The remaining sections of this paper review the structure of the course and assessment; the online environment used for delivery; the characteristics of the students enrolled and reflections of the teaching staff and development team.

Background and Structure of the Programme

The course focuses on the knowledge and skills to select, apply and evaluate business intelligence and data mining techniques to discover knowledge that can add value to a company. Students gain both an in-depth theoretical understanding as well as practical hands-on experience, including implementing novel and emerging techniques. Participants are being kept abreast of current research and state of the art in data analytics related topics. The two year part-time postgraduate programme is delivered over four semesters and comprises of six taught modules (two per semester) and a dissertation. Course structure was influenced by the CRISP-DM methodology.

Technologies to support online synchronous learning

Initial market research suggested a demand for a programme delivered online using synchronous learning technologies. The advantages of online synchronous learning are well documented. Authors such as Collis (1996) and Salmon (2000) outline advantages such as an increase in learner motivation, an ability to deliver timely and quality feedback and a rich online presence which assists in the formation of bonds with fellow learners and faculty. Virtual Classroom tools are a useful component in the facilitation of asynchronous and synchronous learning experiences. During programme design, several virtual classroom systems were considered, namely Adobe Connect,

Microsoft Live Meeting and Elluminate. Adobe Connect was chosen for the delivery platform for a number of reasons:

1. Proliferation of supporting plugin: On the client side, Adobe Connect is powered by the Adobe Flash plugin. The use of a commonly installed plugin such as Adobe Flash can reduce the level of technical support calls from new users. While a user new user to Adobe Connect can make use of the Flash plugin, a new user of Elluminate must have Java installed which may require technical support or administration rights over the client environment. While the Adobe Flash plugin is no longer supported on mobile platforms, access to adobe connect classrooms is facilitated via use of the Adobe Connect App which is available for IOS and Android.
2. Ease of use for participants: Internal trials among ITB academics pointed to a strong preference for the Adobe Connect interface. Customizable Pods and layouts allow meeting hosts and presenters to quickly and easily configure the classroom to the task at hand. This ease of use results in:
 - a. a reduction in the amount of time taken to train academic staff in the use of the tool.
 - b. fewer support queries from academic users.
 - c. a reduction in the amount of time taken for students to become proficient in the use of the virtual classroom.
 - d. a lessening on the reliance of face to face induction sessions which introduce the technology to new students.
3. Feature Set: Adobe connect has a rich feature set including
 - a. The capability to record sessions for achromous learning and study
 - b. VoIP which removes the need to use third party voice systems
 - c. A text based chat tool
 - d. Simple polling and optional broadcast of poll results
 - e. Collaborative whiteboard tool
 - f. Breakout rooms in which students can be grouped
 - g. File sharing
 - h. Screen sharing and the ability to control a remote desktop
 - i. Ability to promote users to different user roles and to grant access to broadcast audio and video to the room
 - j. Reconfigurable layouts and pods

Access to the virtual classroom is controlled via the use of Moodle virtual learning environment (VLE). Authenticated users can access a link to the virtual classroom which is available from within the Moodle course page. Links to recordings are posted which allow participants to review live sessions in the event that they were unable to attend or simply wish to view the session again. Feedback from students highlighted the value of both live lectures and access to recordings for the duration of the course.

The VLE offers a number of key features which are used to create a supportive learning environment for the learner cohort. These include

1. All course notes and resources including links to recorded lectures are posted on the course page
2. Learners can engage in asynchronous discussions through the use of facilitated forums
3. Assignments are submitted via Moodle to TurnItIn using a simple moodle integration. TurnItIn is an Internet based plagiarism prevention service, and integrates seamlessly with assessment submission via Moodle (turnitit.com). TurnItIn allows students check their work for plagiarized content prior to submission, and highlights plagiarized content in the final submission to lecturers. While instances of plagiarism are less frequent at postgraduate level compared to undergraduate level, TurnItIn is a useful service in training students on what constitutes plagiarised content.
4. Grades and feedback (including response files) are posted to each student via the assignment tool.
5. Course calendars inform students of key dates and milestones.

Students

Given the nature of the course, students enroll from wide-ranging backgrounds and geographic locations. As you would expect, the course has attracted students from the local catchment area but also from the wider national and international spheres with international students accounting for 19.5% of the participants between 2011 and 2013 (Figure 1).

The demographics of students enrolled fall into a number of categories:

1. Typically 90% have primary degrees in computing and IT, other disciplines included electronic engineering, pharmacy, forestry management, architecture and GIS.
2. Approximately 15% hold a MSc or another post-graduate qualification
3. Students were aged between 26 and 58, (mean=37±7).

The key characteristic that unites students that enroll in this course is either an existing role that is linked to the data analytics / data management field or students who recognize this course as a unique opportunity to upskill themselves in a field that is ever diversifying.

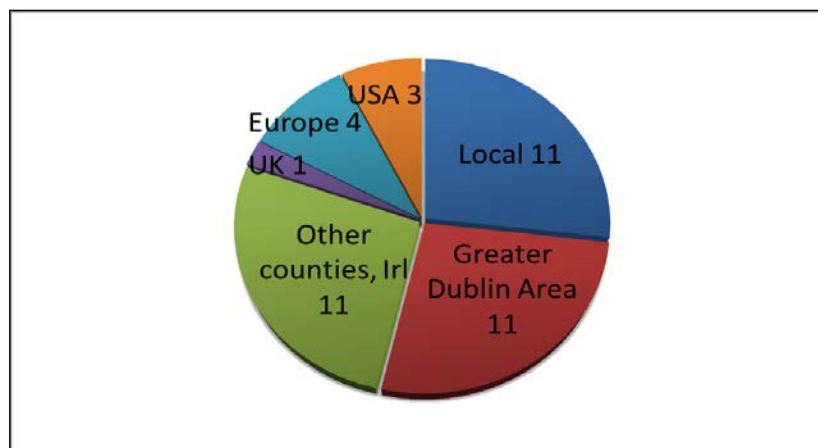


Figure 1. Enrolments by geographic location

Changes to Student Inductions and Supports

Many of the established part time and lifelong learning courses at ITB offered scheduled on site student inductions during which the student is introduced to the institute, facilities, faculty members and support services. However the diverse geographic location of some of the students on the programme meant that not all could not attend the institute for a traditional induction. This led to several immediate issues which needed to be overcome by the programme team:

1. Existing induction procedures were designed on the basis that students would attend campus. For example students were photographed on site for student ID cards and were also shown how to initialise their network accounts for access to campus based IT systems and remote services such as email and moodle. No procedures existed for students who were unable to attend campus.
2. On site induction evenings were a traditional forum for meet and greet sessions with faculty and fellow students.
3. IT and student supports tended to primarily operate during business hours and were not set up to deal with remote students.
4. Services such as the IT helpdesk system were only available to on campus users via the internal intranet.

As a result of the challenges posed by this course the institute initiated a number of solutions which would be of benefit to off campus and on campus students alike:

1. A student information desk (SID) was opened in the library which could deal with face to face queries. The desk had extended opening hours was also equipped with a dedicated telephone number and email address for students who could not attend campus.
2. Weekend and late night IT support was put in place to support delivery of synchronous lectures.
3. A password recovery system was initiated which enabled students to remotely change or recover their network password.
4. A facility was developed to send students their ITB username and a temporary password.

5. Information typically covered during on campus student induction has been incorporated into a detailed student handbook, and sent to all students in advance of the first lecture. This covered access to the online learning environment, managing passwords remotely, interpreting academic grades and GPA, and contact details for course personnel.
6. Students also submit a short bio, photo and contact details prior to commencing. This is compiled into a 'class bios' document each year and circulated to the class.

Assessment of the Students

QQI describe a key learning outcome of a Masters programme as an ability to independently acquire and assess knowledge in novel and emerging technologies. This is particularly true of the rapidly evolving areas of big data and data analytics. We also believe that at master's level, assessment work should enable students specialize in areas of particular interest and relevance to them. Students are assessed using a variety of methods including Research Papers, Practical Reports, Self-evaluations, Peer evaluations, Online Presentations and Dissertation. While guidelines are given on the general area of focus for each submission, students can opt to focus on specific areas and algorithms of interest within those guidelines. Lecture material enables students understand published work, following which student submissions must include greater depth and analysis than is covered in taught content. This assessment approach has resulted in a very high standard of submission from students, examples of which are available at the course website (dataminingmasters). A number of students have published their submissions in peer reviewed journals.

Virtual Learning Environment Statistics

Course delivery is designed to facilitate students who may not be able to attend live lecture sessions, by making recordings available immediately after the lecture takes place. One risk of this approach is possible low attendance at live sessions, increasing the challenge of facilitating student engagement in an online learning environment (Dahalan et al, 2011). In contrast to expectations, we have found attendance has been high at all live sessions. Figure 2 shows the distribution of attendance rates for each student tracked over three student intakes (n=48). 60% of students attended at least 80% of all live sessions.

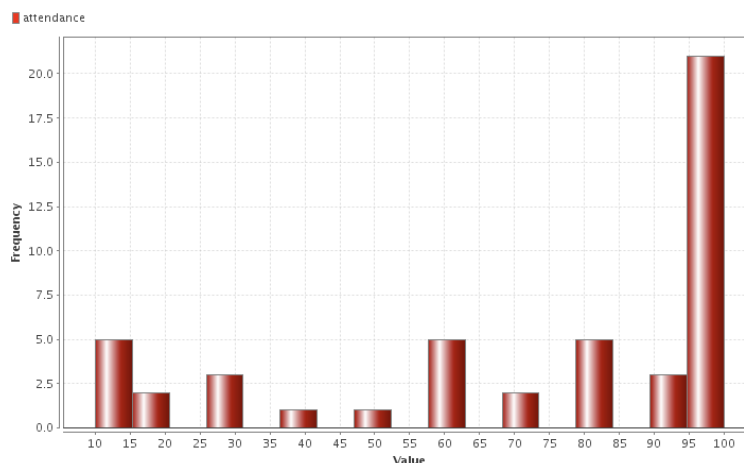


Figure 2 Attendance rate (% of classes attended)

Resources available to students include written lecture, lab sheets covering the practical use of analytics tools, lecture recordings that would incorporate both theory and practical work, research papers, and additional resources expanding on content covered in lectures. Figure 4 illustrates the resources accessed by 2 groups of students (n=25) across two modules. Lecture recordings being the most frequently accessed resource. Most resource requests took place in the daytime, requests in the evening was also high, with a smaller number of requests during night-time hours, as illustrated in Figure 3. Requests represent views to a total of 223 resources, with a number of resources being requested more than one.

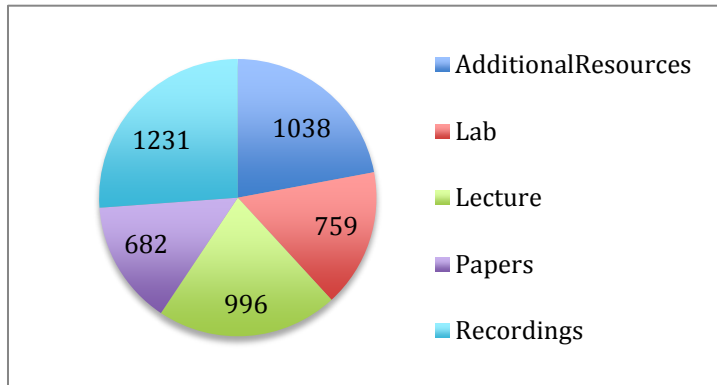


Figure 4. Moodle resources accessed by students

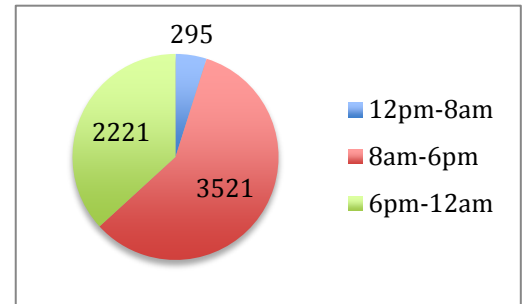


Figure 3. Times resources are accessed

Students are surveyed at the end of semester one for feedback on both academic content of the course, and the quality of the online learning environment (n=28 respondents). All students found the online environment easy to use, found it easy to ask questions and were happy with the level of class interaction. Image and sound quality was good when the lecturer was speaking, but 18% of students had issues with the sound quality when fellow students were speaking. This was generally resolved by the learner using a headset with a built in microphone to communicate with the room. The most valued aspect of the online environment was the facility to record lectures. A recognised challenge of online delivery is harnessing class collaboration (Thorpe, 2002). A minority of students would have welcomed more opportunities to collaborate, 26% of students stated they would have liked greater opportunities for collaboration during scheduled class time, and 28% stated they would welcome more collaboration opportunities outside of scheduled class time.

Regarding course content, all students were happy with the course content, and 96% of students were happy with their quality of learning, pace of delivery, the balance achieved between theoretical and practical work, and the usefulness of continuous assessment in enabling learning.

Reflections

An online learning environment is well suited to the delivery of a data analytics course, particularly at post graduate level. With many high quality open source tools available in this domain, such as Rapidminer (Hofmann & Klinkenberg, 2013), and R (Fox, 2009), practical work is easily facilitated in an

online environment. Datasets are made available to students in advance, and because students have easy access to the tools being used, demos and lab instructions are straightforward to follow. Adobe connect also facilitates student screen sharing, and allows lecturers to request control of a student's screen, again facilitating a physical lab environment.

An important aspect of our programme has been an adherence across modules to a data mining methodology, CRISP-DM (Wirth & Jochen Hipp, 2000). While course content covers the theoretical depth required at level 9, CRISP-DM keeps the focus applied in nature, ensuring that all practical work is done in the context of a business objective to obtain actionable results. This applied focus has both added value to our programme, and resulted in positive feedback on the focus of the programme as a whole.

The course coordinator met with each student individually (online) around week six of semester one. The course also incorporates two modules that facilitate one-to-one engagement with students and their supervisor, a project module in semester two and the dissertation module in semester four. Apart from the academic benefits of these modules, we have found early opportunities to get to know students individually are very valuable in the absence of a traditional classroom setting. We also found that small class sizes (less than 20) help nurture a sense of class belonging.

Postgraduate students tend to have an existing interest in data analytics, and many are already working with data. The nature of data analytics, in that it can be applied to a wide variety of domains, means students engage easily and quickly with techniques, particularly if selecting their own datasets to which they can apply existing domain expertise or interest. We have found it beneficial to use well-cited datasets in semester one, chosen to illustrate the variety of data types typically found in industry, and giving students experience at working with a variety of data types. From semester two onwards, students are free to work on datasets of their choice.

Conclusion

Student feedback indicates that this method of course delivery, the content and value derived and the software vehicle by which it is taught is extremely successful. The course has attracted students both nationally and internationally, and while it is recognised that physical classroom led courses will always be perceived to be the most productive, the positive feedback and results seen strongly indicate that the area of data analytics is suited to this mechanism of teaching and delivery.

Further valuable feedback from students shows that for those working in a day to day data analytics environment, the practical project element of the course permits them to use “real” data and help solve business led data issues in real time.

As articulated in the introduction, in an age where data is king, a course that teaches students (and by association industry) to master and harness its potential, but couples this with the facilitation of this education in a

flexible and feasible manner, is an attractive and successful course. All our feedback points to this being the case.

References

- Collis, B. (1996). *Tele-learning in a digital world: The future of distance learning*. London. International Thompson Computer Press.
- Salmon, G. (2000). *E-moderating: The key to teaching and learning online*. London: Kogan Page.
- Data Mining Masters. (n.d). *Data Mining Masters*. Retrieved January 15th, 2015, from <http://www.dataminingmasters.com>
- John Fox. Aspects of the social organisation and trajectory of the r project. *The R Journal*, 1:5–13, 2009.
- Markus Hofmann and Ralf Klinkenberg, editors. *RapidMiner: Data Mining Use Cases and Business Analytics Applications*. Chapman and Hall/CRC 2013.
- National Framework of Qualification. www.nfq.ie, last accessed Jun 2013.
- Norziani Dahalan Omar, Hasawati Hassan, and Hanafi Atan. Student engagement in online learning: Learners attitude toward e-mentoring. *Procedia Social and Behavioural Sciences*, 67, 2011.
- Mary Thorpe. Rethinking learner support: The challenge of collaborative online learning. *Open Learning: The Journal of Open, Distance and e-Learning*, 17(2):105–119, 2002.
- QQI Award Standards - Computing. (n.d.). Retrieved January 12, 2015, from <http://www.qqi.ie/Publications/Computing - QQI Awards Standards.pdf>
- Rüdiger Wirth and Jochen Hipp. Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, pages 29–39. Citeseer, 2000.