

FlipAround

MDSI student community

Contents

1	Welcome	5
1.1	Checklist of things to do	5
1.2	Pre-flight tests	6
2	The data science mindset	7
2.1	CLARA	7
2.2	Data Science Professional Competencies	8
2.3	Ethics and Privacy	8
2.4	Digital Footprint	9
2.5	Opportunity for overseas exchange	9
2.6	Electives	9
3	A ‘survival guide’ to MDSI	11
3.1	First steps	11
3.2	Technology	12
3.3	Writing	14

Chapter 1

Welcome

MDSI is unique in its approach and feel. MDSI is a ‘boutique degree’ which means we are a small tight-knit data family which means the contacts you walk out (really) knowing are going to be more valuable than the skills you learn. In terms of content, our point of difference is the innovation in our name. We take our innovation component as seriously as data science, and is ingrained in everything that’s taught. For us, a data science degree was our innovation (we were the first of its kind in Australia), and in the rapidly changing context that is data, the ability to innovate and adapt is a pretty great point of difference for you too.

Data science is a collaborative discipline. Students in the MDSI program get hands on experience of working in teams to formulate and solve real-life data science problems. Most courses focus on techniques to solve problems, but spend very little time (if any) on how problems should be formulated. The MDSI program is structured in a way that helps students learn this tacit, but crucial skill.

Another important aspect of data science is that it is a rapidly evolving field. A data scientist must therefore be able to stay current with developments in the field. The MDSI program, with its emphasis on critical self-learning, prepares students to be lifelong learners.

Welcome, and good luck on your MDSI journey

1.1 Checklist of things to do

Getting started on your MDSI journey can be somewhat overwhelming. So to help you ease into life as an MDSI student, the following checklist will help you to get up and running as painlessly as possible.

- Do your statistics pre-flight test
- Activate UTS student email
- Forwarded UTS student email if required
- Review Subject Outlines
- Activate and personalise CICAround
- Do your CLARA profile
- Log into Diigo
- Join the Slack Channel
- Log into Review
- Log into SPARK
- Test your Google Drive
- Test your Office 365 Drive
- Download R & R Studio
- Download Python & Rodeo

- Download Tableau
- Download KNIME

1.2 Pre-flight tests

MDSI statistics pre-flight test: <http://www.uts.edu.au/future-students/analytics-and-data-science/essential-information/mdsi-statistics-pre-flight-test>

Chapter 2

The data science mindset

2.1 CLARA

Each person has their own learning preferences and habits of mind that shape their response to challenges and learning opportunities.

CLARA (Crick LeArning for Resilient Agency) is a tool used to prompt reflection on a multidimensional construct called “Learning Power” with eight dimensions: curiosity, creativity, sense making, belonging, collaboration, hope and optimism, mindful agency and openness to change. The UTS Graduate Attributes have a strong resonance with these dimensions. The CLARA tool is used as part of MDSI activities, aiming to help the students maximise their development results through understanding themselves better, namely with regards to their approach to learning and challenges.

The tool is survey-based and provides a profile-style feedback covering each of the following dimensions: Curiosity - Wanting to get beneath the surface & find out more - Always wondering why and how Creativity - Using my intuition & imagination to generate new ideas & knowledge - Taking risks & playing with ideas and artefacts to arrive at new solutions Sense making - Making connections between what I already know & new information & experience - Making meaning by linking my story, my new learning & my purpose Belonging - Being part of a learning community at work, at home, in education & in my social networks - Knowing I have social resources to draw on when I need them Collaboration - Being able to work with others, to collaborate and co-generate new ideas and artefacts - Being able to listen and contribute productively to a team Hope and optimism - Having the optimism & hope that I can learn & achieve over time - Having a growth mindset; believing I can generate my own new knowledge for what I need to achieve Mindful agency - Taking responsibility for my own learning over time through defining my purposes, understanding and managing my feelings, - Knowing how I go about learning & planning my learning journey carefully Openness to change - An emotional orientation of being open & ready to invest in learning, having flexible self-belief, willing to persist & manage any self-doubt - A necessary prerequisite for developing resilience in learning

Here is an example of an output from CLARA, showing the resulting profile, based on the responses provided in the survey.

Insert images here

Learning profiles can change over time, so MDSI conducts two sets of CLARA tests, to allow the students to compare their profile changes and reflect on their development progress.

CLARA survey will be organised by MDSI and each student will receive a link and instructions on how to utilise this tool.

2.2 Data Science Professional Competencies

Competency can be defined as “effective application of skill, knowledge and abilities to on-the-job-behaviour and capability to perform to job requirements”. MDSI supports the ongoing development professional development efforts and offers a tool that can help the students evaluate their skills and abilities in various domains associated with the data science professional competencies. Students are encouraged to utilise the tool to identify the key competencies associated with their individual career aspirations. For each relevant competency, students evaluate their current competency levels, identify any gaps and use the information to create a set of actions that would form their professional development plan. The competencies model is introduced to students as part of 36100 (Data Science and Innovation) subject through a self-assessment exercise. The competencies are divided into two groups: technical and managerial, describing the following competency (proficiency) levels for each domain: Beginner, Competent, Advanced and Expert.

Technical: - Mathematics and statistics - Programing and scripting languages - Databases and data storage - Computing systems, platforms, security, integration - Data mining - Data modelling - Analytics, predictive modelling and machine learning - Data visualisation - Business analysis and interpretation - Product development

Interpersonal/managerial: - Creativity - Communication - Data strategy - Line management - Data management and governance - Facilitation and presentation - Project management

Competency levels: **Beginner:** able to assist and perform simple tasks **Competent:** able to perform tasks independently **Advanced:** able to perform complex tasks **Expert:** able to perform complex transformative, strategic or trans-disciplinary tasks

The competencies assessment exercise aims to prompt the students to take a proactive attitude to their professional development efforts and effectively apply their analytical skills, dedication and professionalism in managing their career objectives.

The competencies assessment exercise covers the following steps: 1. Evaluate your current competency level for each domain on the list 2. Choose a set of domains (no more than 6) that are relevant to your planned development for this subject, your course and your career goals. 3. Identify the goal competency levels for the selected domains and describe related professional development outcomes that support your assessment 4. Analyse your development outcomes in the context of your career goals and identify the gaps between your current and goal competency levels 5. Develop a set of actions needed to achieve desired level of competencies and bridge the identified gap 6. Provide feedback and suggestions for the improvement of the current list of competencies, descriptions etc.

2.3 Ethics amd Privacy

Its important to understand that security, privacy and ethics are three different things, although heavily intertwined in the ‘internet of things’.

What is ethical when it comes to data and the internet of things? Is privacy having a login or not being identifiable as an individual?

The world of Ethics and Privacy is changing, similar to the definition that now includes much more than it did a decade ago. Computer security like a login is no longer sufficient to providing protection of privacy which is more focused on ensuring that only people who should have the authority to access your information should be able to.

Current Privacy legislation addresses control and authentication processes of whom can access your information via direct disclosures and how this information should be stored by the party who is collecting this information, it does not address disclosures that can be made based on inferences that can be drawn from big data of which your information is a part. Is the value or conclusions that could be drawn from information as part of big data considered private information?

A sensible framework in relation to Ethics and Privacy where data is concerned is highlighted in the Belmont report which identifies two rules to consider “(1) do not harm and (2) maximize possible benefits and minimize possible harms.”

A big ethical dilemma of late is the rich data sources that various provider hold, that if pooled together will strip all possibility of anonymity.

For more on this read:

<http://www.tandfonline.com/doi/full/10.1080/08900523.2014.863126?src=recsys> http://libres.uncg.edu/ir/uncg/f/N_Kshetri_Big_2014.pdf

2.4 Digital Footprint

Your digital footprint is the name given to the data that is recorded about you all day every day. It can be the time and phone number of someone that you called, the mobile phone tower that you were connected to at the time of making the call and how long you spoke for. It is the IP address of your computer when you connect to the internet. It is the list of items you pay for when you go through the checkout at the supermarket and the eftpos card number you used to pay for the items. It is the surveillance footage you appear in when you move through monitored public spaces. It is stories you ‘like’ or share on social media sites. It is the journeys that your GPS navigation stores about your travels. It is every email you send and every click you make when you browse the internet.

Your digital footprint is the inescapable record of your existence by doing nothing more than living your life. It is an important aspect of modern society as many services that you enjoy depend on the data you generate in order to provide critical services. A bank can’t tell you how much money you have without keeping record of your bank transactions. For good or for evil, this data comes embedded with far more information about you. By looking at the kinds of things you spend your money on or the businesses that you spend your money at and the time of day that you spend your money there, it can be determined where you live and where you work.

As an MDSI student, you will learn to think critically and ethically about data collection and how it can be used for good and for evil. The best place to start your thinking is with your own digital footprint, become aware of how big it is and how you feel about it.

It’s important to note that very little permission is sought on data collection and when it is sought, very little education is provided in terms of the use of that data. Very few providers who collect data clarify or specify what the data they collect is used for.

You are responsible for your digital footprint. Generate it wisely.

2.5 Opportunity for overseas exchange

Some great opportunities exist within MDSI with our Program Director having many contacts in many other countries which enable us to be able to explore greater opportunities for overseas exchange.

You need to do a few things before this opportunity is explored as set out by the Program Director to ensure for an easier way forward if this is an opportunity you want to explore.

2.6 Electives

You need to select four electives during your MDSI course. These electives should be selected to assist you in your growth as a student and as a data science professional. These subjects enable you to add to your

toolbox of where you are heading with your journey.

Electives can be selected from any school however you will still be subjected to the pre-requisites for any possible subject, so it will depend on the requirements of the subject.

We suggest that when you apply for a subject with a prerequisite that you also apply for a waive of the prerequisite if the prerequisite is a subject you are familiar with but have not done with UTS and get exemption for that requisite.

This is not always easy, or approved and is subject to each School's internal views or policies. It is definitely a consideration to take.

You can apply for the subject (and a waiver of prerequisites if required) early as CIC is not limited by inter-faculty time restrictions.

Our best tip is : get in early.

Chapter 3

A ‘survival guide’ to MDSI

3.1 First steps

3.1.1 Your UTS email:

First and foremost you need to activate your UTS email address. All official communications from UTS, subject notifications, MDSI newsletters etc will be sent to this email address. You need to activate your email address before you can access other UTS systems.

Activate your UTS Student email:

- Navigate to <https://email.itd.uts.edu.au/webapps/myaccount/activation/> and follow the steps to activate your UTS student email account.
- **Protip:** If you don't want to login frequently to check if you have mail, simply setup a email forwarding to an email address of your choice via the settings page after logging in.

3.1.2 Get your subject outlines:

MDSI uses a variety of systems for online teaching and learning. UTSONline and CICAround are the two primary environments for you to familiarise yourself with.

The first thing you need to do after activating your email address is to login to UTSONline, access your subjects and find your subject outline. Your subject outline contains everything you need to know about your subject for the coming semester. It includes the contact information for your subject co-ordinator, important dates, assessment descriptions and much more. In most cases you can find the answer to any question you might have about your subject addressed in the subject outline.

Find your subject outlines in UTSONline:

- Login to UTSONline at <https://online.uts.edu.au/> using your student ID number and the password you setup for your UTS email account.
- Access your subjects by clicking on your subject name
- Download your subject outline by clicking on the link titled ‘Subject Outline’ in the left side menu, then click on the subject outline link on the page.

3.2 Technology

3.2.1 Join the MDSI community:

Your next stop should be CICAround. Here you will connect with your peers in an academic capacity. There are discussion forums for your subjects where you can post questions. CICAround most notably is where you will go to blog about your experiences throughout your MDSI journey. The first step is to activate your blog. Then you can browse through the blogs of your new MDSI family and read about their experiences and the things they have learnt.

Activate and personalise your CICAround profile: - Navigate to <https://ca.uts.edu.au/using-ca/> - Watch the welcome video then login to CICAround using your student ID and password. - Put up your first CICAround Blog post

3.2.2 Join the MDSI chatter:

Slack has proven to be a very useful tool so far. It is completely driven by the student community and is where the MDSI student community goes to socialise, organise BBQs, ask each other for technical help and share useful resources. If you need a quick answer, Slack is the place to go.

Join the Slack Channel - You can download the Slack application from <https://slack.com/downloads> - You can also get the app for IOS, Android and Windows phones. - You do not need to pay for a subscription. - Signup to the MDSI group at: <https://utsmds.slack.com/>

3.2.3 Diigo

A collection of resources contributed by the MDSI community.

Join the Diigo group - simply create a Diigo account and request access. https://groups.diigo.com/group/cic_mdsi

Frequently used search tags include: > “*DSI,DAM,Data,big_data,case,studies,visualization,teaching_tools,statistics,stats-thnkg,privacy,Algorithms,ethicsVSD,realworldDM,video,Analytics,human-machine,history,data,mining,Data_science,cisco,r,sets,accountability,digital_futures,tools,DM,reading,DVN,equality,infographic*” >

3.2.4 Review

Review is an assessment tool that is used to mark your work, give you feedback about your work and for you to develop a sense of what is expected by marking your own work before it is assessed by teaching staff.

3.2.5 SPARK

SPARK is an acronym for Self and Peer Assessment Review Kit. This tool has been developed to assist with in class activities as well as being able to self and peer assess assignments. Its also one of the few tools that make group assessments/work easier, in particular to marking.

3.2.6 Google / Office 365

Your university account allows you access to Google Apps and Office 365.

Google Apps access does not include Gmail. You can not login to your UTS google apps account via gmail or if you are already logged in with a personal gmail account, you will need to log out completely from gmail.

Once you have done this, you can log in using your student email address. This will revert you to a UTS login page. Use your UTS student number and password and it will revert you back to the Google Drive, but you will be logged into the drive.

Similarly you can mimic the same steps for Office 365.

3.2.7 R / R Studio

‘R’ is a coding language used by most of the data science community. RStudio is a software program or ‘Integrated Development Environment’ (IDE) that makes working with the R language ALOT easier. The programming environment is really flexible as it allows you the joy of working in a notebook format, scripting, markdown and publishing your work as a PDF.

You will use R in many of your subjects and being able to use it well will give you a serious edge over your classmates and competitors at hackathons.

Download and install R & RStudio free - Download and install the R language: <https://cran.rstudio.com/> - Download and install RStudio IDE: <https://www.rstudio.com/products/rstudio/download/>

Libraries well worth their weight in gold: - tidyverse - <http://tidyverse.org/> A collection of libraries that make data analysis easier - readr - <http://readr.tidyverse.org/> for reading all kinds of data formats - stringr - <http://stringr.tidyverse.org/> for working with text - ggplot2 - <http://ggplot2.tidyverse.org/> for visualising data - tidyr - <http://tidyr.tidyverse.org/> for creating tidy data - dplyr - <https://github.com/hadley/dplyr> for manipulating data - caret - <http://topepo.github.io/caret/index.html> for creating predictive models - Bookdown - <https://bookdown.org/> for creating beautiful documents

There are many resources to get you started in doing data science with R. Refer to the resources section for more information.

3.2.8 Python / Rodeo / Jupyter Notebook

Python is a general purpose coding language widely used by the data science community. A great place to start is with the Anaconda from Continuum Analytics : <https://www.continuum.io/downloads>

Python for data science is commonly used in a notebook format. To this end Jupyter notebooks will become a familiar friend. Fortunately it is included as part of Anaconda. For more info, refer to the resources section.

If you prefer an ‘R’ style IDE, then Rodeo by Yhat is for you. <https://www.yhat.com/products/rodeo>

If you prefer a traditional IDE, you can get a free license for PyCharm (as well as all their other products) from JetBrains using your student details: <https://www.jetbrains.com/student/>

Libraries well worth their weight in gold: - Numpy - <http://www.numpy.org/> for working with numerical arrays - Scipy - <https://www.scipy.org/scipylib/index.html> for scientific computing with python - Matplotlib - <http://matplotlib.org/> for visualising data - Seaborn - <http://seaborn.pydata.org/index.html> for statistical visualisation - Pandas - <http://pandas.pydata.org/> for working with data - Statsmodels - <http://www.statsmodels.org/stable/index.html> for creating statistical models - Scikit-Learn - <http://scikit-learn.org/stable/> for doing machine learning with python - Tensorflow - <https://www.tensorflow.org/> ‘deep learning’ with python

Note: Python comes in two different flavours: 2.7 and 3.x. You can start with either version, but it is worth learning what the subtle differences are (eventually).

A couple of blog posts to help you choose between R and Python: - R vs Python for Data Science: The Winner is <http://www.kdnuggets.com/2015/05/r-vs-python-data-science.html> - R vs Python for Data Science: Summary of Modern Advances <https://elitedatascience.com/r-vs-python-for-data-science>

3.2.9 Tableau

Tableau is a tool for visualising data. It is quite powerful in its ability to connect to a variety of data sources both on your computer and through the internet. It is also relatively intuitive to use.

As a student you can apply to the company for a free license to their commercial desktop version. <https://www.tableau.com/academic/students>

3.2.10 KNIME

KNIME Analytics Platform is an open source solution that enables quick, fast data driven designs for machine learning. Its a visual tool to learn and use when you need to get the job done quickly (without writing any code) and need to create algorithms quickly but you don't have the time to learn the mathematics behind the algorithms. Its friendly and easy to use to find the hidden 'story' in the data.

Go to <https://www.knime.org/knime-analytics-platform> to download KNIME for free.

3.2.11 Rapidminer

Rapidminer is another visual tool for doing data analysis, modelling and machine learning. You can get access to their commercial tools using your student status from <https://rapidminer.com/educational-program/>

3.2.12 QGIS

QGIS is a really nice open source tool for working with geospatial data. To get started just head over to <http://www.qgis.org/en/site/index.html>

3.2.13 Github

Github is a fantastic tool to get familiar with. It is a great place to store code, collaborate with other and even host your own website or blog.

Github has a really generous collection of free stuff for students. To claim yours head over to: <https://education.github.com/pack>

3.3 Writing

3.3.1 Blogs

More and more academics and workplaces use blog posts to reach clients, audiences and share knowledge. Blogs can be useful for many reasons and is used as a reflective tool for students as well as providing an opportunity to share any learning.

You can use some tools to turn topics into amazing titles by using keyword suggesters (<http://keywordtool.io>), title generators (<https://www.portent.com/tools/title-maker>), and you can also test your headlines with the following tool (<http://coschedule.com/headline-analyzer#>)

Tips for new bloggers - Use an eye catching title - In-text links - Use pictures, pictures speak a thousand words - Keep post to 1000-1500 words - Use social sharing buttons - Use paragraphs - one idea per paragraph - Revise and Rewrite - Omit needless words - Use the KISS (Keep It Simple, Stupid) Principle - Use definite, specific concrete language - direct and to the point - Write in a way that comes naturally - use your active

voice - Be clear - make it simple to read and understand - Avoid fancy words - Do not take shortcuts at the cost of clarity

Tips on writing blog posts : <https://probblogger.com/how-to-write-great-blog-content/> and <http://www.socialmediaexaminer.com/26-tips-for-writing-great-blog-posts/>

3.3.2 White papers

Where do you start with a white paper and what are they? White papers are originally documents written for government policies however this is most recently being used by companies and universities to get new policies and research into the public space.

There are some things to consider when writing a white paper: - Pick a topic people will want to read or a problem you want to solve - Pick a generic title that describes the problem at hand - Engage, inform and convince your reader - Be descriptive and professional - Consider the audience you are 'speaking' to and accommodate for their level of expertise - Set up a great intro - Emphasize the value you want to or will create - Decide on a length for the white paper (1-5 pages are the norm) - Describe the solution you are proposing - Remember a summary that reviews the problem, solution and result of the outcome - Proofread your document, and ensure someone else reads it before you submit/publish it. - Follow the 3-30-3 rule (you have three seconds to captures your audience's attention from a glance at your piece, if you succeed at capturing their attention then you have 30 more seconds to ensure they continue reading, if you pass the 3-30 rules then your reader will give you three more minutes to make your point).

3.3.3 Assignments

You will get an assignment brief for each assignment you have due. Its is recommended that you review these briefs and that you follow the detailed instructions set out for you.

Bibliography