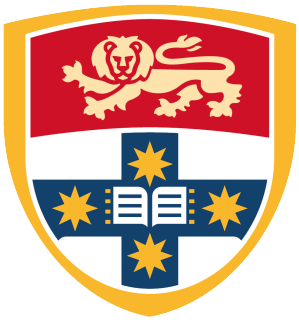
INFO1111 - Group 46

Bachelor of Advanced Computing Computational Data Science Major

AYESHA MAYAT

SID: 430345827



THE UNIVERSITY OF SYDNEY

Tutorial ID: F10A

School of Computer Science
Faculty of Engineering and IT
The University of Sydney
Australia

21 June 2019

CONTENTS

Chapter 1 Overview of the major **3**
1.1 Computational data science 3

Chapter 2 Overview of the timeline and courses of the major **4**
2.1 Courses of major 5
2.2 Evaluation of interesting courses 6

Chapter 3 Expected career path **7**
3.1 Types of roles available 7
3.2 Skills and tools required for these roles 8
3.3 Similarities and differences between majors 9
3.4 Companies hiring graduates of major 10
3.5 Careers fair observations 11

Chapter 4 Emerging Possibilities **12**
4.1 Industry progression 12
4.2 Potential problems of the future 13

References **14**

CHAPTER 1

Overview of the major

1.1 Computational data science

The Computational Data Science (CDS) major in the Bachelor of Advanced Computing is a four year degree with the option to do one of Information Systems, Computer Science, or Software Development as a second major [**7**]. This paper will focus on the CDS major as a stand alone major.

An individual with a CDS major has skills that are sought after in the engineering, science, business, and many other industries where data is collected daily. The CDS major also teaches the value of this data and the tools to utilise it for improvement.

The CDS major develops both technical and soft skills needed in the IT industry. The technical skills include programming, analysis, and data-driven decision making which are key for a competent professional. In addition, the soft skills encompass team work, communication, ethical behaviour, academic integrity, and professional development. These skills are highly valued in the work force with more focus on them in recent years.

Graduates can develop a career in a new and rapidly evolving field of IT covering machine learning, artificial intelligence systems, and predictive data analysis. The improvements in both hardware and software make studying the CDS major an exciting path of analytical thinking and innovation.

CHAPTER 2

Overview of the timeline and courses of the major

The Bachelor of Advanced Computing with a major in Computational Data Science runs over four years and requires a total of 192 credit points for completion [**7**]. This comprises of 96 credit points of core study units, 48 credit points towards an approved IT major, and at least 12 credit points of IT electives at the 4000 level or higher. Additionally, students have the option to choose up to 12 credit points in the Open Learning Environment, a minor of 36 credit points or a second major of 48 credit points, and electives where appropriate. The available minors and second majors are Computer Science, Information Systems, and Software Development. Honours is awarded to students who achieve Honours Class I: 75.

Students can also choose to pursue the Dalyell stream which requires admission and completion of 12 credit points of Dalyell units and a maintained WAM as determined by the Board of Interdisciplinary Studies.

2.1 Courses of major

Year	Semester 1	Semester 2
1	MATH1021 (Calculus of one variable) MATH1002 (Linear algebra) DATA1001 (Foundations of Data Science) INFO1110 (Introduction to Programming) INFO1111 (Programming Professionalism)	MATH1064 (Discrete maths) ELEC1601 (Introduction to Computer Systems) INFO1112 (OS & Network platforms) INFO1113 (Object-Oriented Programming)
2	DATA2001 (Data Science: Big Data and Data Diversity) COMP2123 (Data Structures & Algorithms) INFO2222 (Usability & Security) Elective	DATA2002 (Data Analytics: Learning from Data) ISYS2120 (Data & Information Management) SOFT2412 (Agile Software Development Practices) Elective
3	DATA3001 (Data Science Capstone Project) INFO3333 (Management) One of: COMP3027 (Algorithm Design) or COMP3308 (Introduction to AI) Elective	One of: DATA3404 Data Science Platforms DATA3406 Human-in-the-Loop Data Analytics Elective Elective Elective
4	INFO4444 (Computing 4 Innovation) INFO4001 (Thesis A) Advanced Elective Elective	INFO4002 (Thesis B) Advanced Elective Elective Elective

NB: Students starting in semester two will study INFO1110 in their first semester and INFO1113 the following year in semester one.

2.2 Evaluation of interesting courses

ELEC1601 - Introduction to Computer Systems : Categorised under electrical engineering, this unit is an introduction to the core concepts behind computer design and operation. It covers the CPU and computer organisation, logic gates, machine and assembly languages, and high level programming constructs.

DATA2001 - Data Science: Big Data and Data Diversity : DATA2001 is where students begin to work with large data sets. It teaches data collection, data cleaning, processing, and querying. The unit utilises Python libraries together with SQL to introduce basic declarative data processing and analysis on real world data.

COMP3308 - Introduction to AI : This unit is an introduction to the concepts behind AI design and functionality. It explores machine learning, neural networks, game playing, probabilistic reasoning, and other AI topics. It also involves a practical component where simple AI problems are solved using the techniques learnt.

INFO4444 - Computing 4 Innovation : Innovation is a significant part of the IT field and INFO4444 focuses on understanding the relationship between business value and innovation through the Business Canvas Model. Students will hear from industry innovators about successful and unsuccessful businesses based on innovative ideas.

INFO4001 and INFO4002 - Thesis : The fourth year of the Bachelor of Advanced Computing is an honours year. Students write a thesis on a topic of their choice with honours awarded to those with sufficient merit.

CHAPTER 3

Expected career path

3.1 Types of roles available

The skills learnt in the CDS major can lead an individual down many different and varied career paths. These range from pure academia to quantitative analysis focusing on the data only to qualitative analysis and consulting for the improvement of business. The major is relevant in all businesses and industries where data is collected. In the modern connected world, this covers every area of work. The roles discussed below are but a few of the many jobs for which a CDS gradutate has the knowledge and capability to pursue.

3.1.1 Data Analyst

A data analyst uses the core technical skills taught in the CDS major to process, query, summarise, and visualise data. They focus on quantitative data results and work to solve given problems using reliable knowledge and trusted techniques. They perform predictive data analysis utilising their knowledge of statistics, data transformation and visualisation, and of course, data analysis. Ultimately, data analysts propose projects for improvement.

3.1.2 Data Scientist

A data scientist can be thought of as an advanced data analyst. They transform raw data into useful information in the same way that data analysts do. However, a data scientist will take their data exploration to the next level. As well as providing the statistical information, a data scientst will explore the data without a given problem to solve. That is, they will explore the data. This requires extensive knowledge of computing techniques and a creative mind. Data scientists are able to work in the private, public, government, and all other sectors.

3.1.3 Business Information Analyst

Individuals who have graduated with an Information Systems major as well as the computational data science major can pursue a career as a business information analyst. A business information analyst evaluates a company's IT system and recommends improvements using gathered data and analysis. They are involved in the process from development to testing with the aim of making the company more efficient. They consult stakeholders to understand the issues at hand in order to develop these solutions. Business information analysts also write project reports and are responsible for the IT systems of the company performing additional analysis where needed.

3.1.4 Data Analyst Consultant

Similar to a business information analyst, a data analyst consultant performs qualitative data analysis as well as quantitative data analysis. This is often under employment from a consulting firm for clients in all sectors of business. The analysis goes beyond the data alone. Data analyst consultants consider the broader scope of the business including the business' mission, management, human and other resources, and its IT systems. They are responsible for developing the techniques and methods to implement approved projects as well as its success. Therefore, they may not delve as deep into the data as a data analyst who purely focuses on quantitative data analysis. Data analyst consultants distinguish between necessary and unnecessary items for the business.

3.1.5 Big Data Developer

Big data developers create the software used by data analysts i.e. Hadoop applications. The job description is similar to a software developer except that they will focus on working around big data. Big data developers require the knowledge of data analysts to understand the problems faced when dealing with big data and the processes to deal with them. In addition, they work on scalability; taking software to the next level with the aim of being able to work with large quantities of data at high speed.

3.1.6 AI Developer

The minds behind intelligent systems are AI developers whose job it is to "teach" the system. An AI developer works to improve or create custom intelligent systems for their company or clients. They can be involved in many stages of the project development from functional specifications and design, development, and integration, to testing and deployment. AI developers require creativity to translate requirements into working intelligent systems.

3.2 Skills and tools required for these roles

3.2.1 Technical Skills

Technical skills encompasses the knowledge and ability to work with the data itself. It includes, but is not limited to:

- SQL (structured query language) is a simple yet powerful tool in any profession handling data. It can write, read, update, and retrieve data through its declarative statements.
- Python is a high level programming language that is compatible with many other languages. Its wide variety of libraries makes it a very versatile language for data processing, manipulation, and analysis.
- Julia is another high level programming language designed for numerical analysis and computation. It can call C functions directly and Python functions through PyCall.
- MATLAB is designed for technical computing. It combines computation and visualisation to produce mathematical constructs direclty.

- TensorFlow is used in machine learning, neural networks, and artificial intelligence systems. It can operate over many platforms including CPUs, GPUs, and TPUs performing computation and outputting nice stateful dataflow graphs.
- SAS (Statistical Analysis Software) performs advanced analytics, multivariate analysis, data management, predictive analysis, and more. It is a software that can retrieve data from multiple sources and statistically analyse it. SAS covers all levels of expertise having both graphical and textual user interfaces.
- JSON (JavaScript Object Notation) is a text format that uses computing language convention which make it idea for data interchange. It is both human and machine readable. JSON is often used inconjuction with Python.
- Scalar
- HTML (HyperText Markup Language) is usually used for designing web browsers. Since the internet is a great resource for data, graduates should be familiar with HTML.

3.2.2 Soft Skills

As well as having the knowledge to work with the data, CDS graduates must also have the skills to be able to work alongside and collaborate effectively with the professionals around them. This includes:

- Communication is essential when working with others in all fields including IT.
- Team work - Most jobs require multiple people to complete the task at hand and so team work is a must. Teams that work well together are more productive and successful.
- Honesty is necessary when determining competency for a task, and when working with clients and colleagues.
- Academic integrity is important as it ensures credit is given where it is due. It means acting with honesty and acknowledging other's work when used in your own.
- Professional development is essential for an IT professional to stay relevant in a rapidly progressing and evolving field. With constant development and improvements comes the need to know the latest technology to work efficiently.

3.3 Similarities and differences between majors

3.3.1 Computer Science

- The major of computer science has a strong focus on computer theory and the application of the theory to create computers and solve computing problems. A computer scientist will ensure a system is working as expected to give the data a data scientist is interested in.
- There is common ground in the field of machine learning and artificial intelligence where data and computer scientists come together to build a system that can learn for itself. The computer scientist has responsibility over how the systems works and the data scientist has responsibility over the learning algorithms of the system and feeding it data for machine learning.

- Another commonality is in the area of business consulting. A data analyst consultant handles the data, and together with a systems analyst of the computer science major (see Computer Science Major report), they bridge the gap to software engineers. Together they are able to combine technical computing knowledge and data-driven decision making to suggest solutions for problems or overall improvements to functionality.

3.3.2 Software Development

- There is a crossover with computational data science where software developers, like previously discussed big data developers, create programs for handling big data (see Software Development report).
- Otherwise, software developers differ from computational data scientists greatly in their work. Data scientists focus on the data and use the programs software developers create.
- The two fields may work together when a data scientist needs a program to run a certain way and thus will consult a software developer to create it. Alternatively, a software developer could consult a data scientist to find out how clients use their software in order to get valuable information that would help them improve their software or develop a software to cater to gap in the market.
- The skills needed overlap for the most part, with many programming languages being common. The majors differ where data scientists require languages that where specifically developed for use with big data processing.

3.4 Companies hiring graduates of major

3.4.1 ResMed

ResMed develops solutions to treat sleep and respiratory disorders [**5**]. Their medical devices utilise technology which allows people with related diseases to be treated in the comfort of their own home. These medical devices are cloud-connected and collect data to provide cutting edge care and service for customers. ResMed has a variety of disciplines graduates can enter from product development, manufacturing and operations, engineering and software development. They can work with world-class teams within the various departments of ResMed's businesses. Additionally, graduates have the opportunity to become forerunners in their career paths creating medical devices with ResMed.

3.4.2 SAS Institute

SAS Institute is a developer of analytics software called SAS; Statistical Analysis System [**6**]. SAS software is used to retrieve, manage, analyse, and create reports on data for data-driven decision making. SAS offers academic programs to ensure their employees are qualified. The company has been listed by Fortune magazine as a "Best Company to Work For" for over twenty years. Through SAS. graduates are able to match their skills with customers' needs and put their knowledge to use.

3.4.3 Infosys

Infosys specialises in digital services and consulting. It helps customers transform their businesses through technology and its AI-powered core. Infosys embraces advancing technology to build intelligent systems with the aim of innovation. The company has helped people in farming, commerce, enterprising and more to realise the potential of their product and businesses through intelligent automation and data-driven decision making. Graduates have the opportunity to work across Australia in areas of software development, program testing, production support, management and documentation.

3.5 Careers fair observations

The careers fair hosted companies from a wide range of backgrounds and expertise. These companies where advertising for graduate programs and jobs, some of which have been described in the section above. The careers fair highlighted the various fields a graduate of computational data science can pursue a career in. This includes companies like ResMed and Cochlear which focus on medical technology, SAS Institute, RecordPoint, and Atlassian which innovate using advanced technology , and Macquarie Cloud Services which increase a business' accessibility.

CHAPTER 4

Emerging Possibilities

4.1 Industry progression

4.1.1 IoT

The Internet of Things (IoT) refers to the connectivity and interaction of devices over the internet [**4**]. There are over 25 billion internet connected devices in 2019 and this number is expected to triple by 2025. The growth of these devices has seen an increase in production with an estimated 2.5 quintillion (i.e. 2,500,000,000,000,000) bytes of data produced each day. Such examples are wearable devices including medical devices which are improving emergency response times. Smart homes contain appliances that can be controlled remotely from similar smart devices. The data collected is being used to enhance productivity and efficiency across a variety of organisations using predictive analysis and the people behind this are data analysts.

4.1.2 Improving Hardware

Along with more devices becoming connected in the IoT, there is a corresponding trend in improved technology for hardware. Just think about how often Apple releases a new iPhone or other smart device. An example of this is Google's Tensor Processing Unit (TPU) [**3**]. This tiny processor is smaller than a five cent coin and has more processing power than graphic processing units (GPUs) and standard central processing units (CPUs). In 2017, Google quoted it to be 27 times faster than GPUs but this number is likely to have increased even within the two years since its public release.

Another example of improvements in hardware are optical transistors. Currently, the standard transistors being used are electronic, made of silicon, and require a voltage to switch signals. The size of these electronic transistors are reaching their physical limit being about 10-20 nanometers and expected to reduce to 5-7 nanometers over the next few years. At this scale the laws of physics get interrupted and it is no longer a question of "Can we make it smaller?" but rather "Will it still work?". Optical transistors have thus been developed to overcome this problem. They utilise light rather than electricity to switch signals and although practically at this stage they are not as small as their predecessors, they have the ability to become so.

4.1.3 Machine Learning and AI

Machine learning is a type of artificial intelligence (AI) system whereby machines are able to improve themselves through data processing and analysis. The field is developing rapidly due to the strides in data collection and hardware mentioned above. Machine learning requires data, the

FIGURE 4.1: Google's AlphaGo mastered the game Go[**1**]

two go hand in hand. Systems like Google's TPU which was designed specifically for TensorFlow, a Google software used for machine learning and artificial intelligence systems, make data processing in AI units faster than it ever was [**3**][**1**]. The large amounts of data available means a single unit can learn a lot in a short period of time and the field of machine learning progresses.

AI has many real-world applications from facial recognition in your Snapchat filter to smart assistants you can talk to like Siri and Alexa. Google's Alpha Go and IBM's Deep Blue have beaten the world champions in Go and chess respectively. As such, AI is proving its worth having been able to do better than the world's best. Its future lies in innovation and proving itself to be economically sound in business. The potential of AI has only begun to be realised and a technology driven world, one even more dependent on technology than we are now, is a very likely scenario.

4.2 Potential problems of the future

The main problem graduated may face in the future is the decline in the skills that are in high demand now [**2**]. As AI and similar technology has improved over previous years, and with automation becoming better, many jobs have been overtaken by technology which is easier and cheaper to use. Examples include self checkout lanes which can be overseen by one employee as opposed to one per counter. This is only the beginning of automation. Through machine learning intelligence systems can already work at speeds way above a human's best. The threat to graduates of the future is when these systems can ask the questions a data analyst would and produce solutions itself. It is difficult to say when this will occur. There is still a long way to go in the relatively new field of machine learning and AI but the speed at which the industry is progressing is great. Analytical and and other high level skills are the skills graduates should focus on developing over others that can, and likely will, be automated in the future.

References

[1] Alpgago. https://qbi.uq.edu.au/files/27603/AlphaGo-Google-AI$_Q BI - 700.jpg$.

[2] Data analytics artificial intelligence: What it means for your business and society. https://www.imd.org/research-knowledge/articles/artificial-intelligence-real-world-impact-on-business-and-society/.

[3] Google tpu. https://cloud.google.com/tpu/.

[4] Iot. https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT.

[5] Resmed. https://www.resmed.com/au/en/index/about-us.html.

[6] Sas institute. https://www.sas.com/en$_a u/home.html$.

[7] Uos outline. https://sydney.edu.au/courses/units-of-study/2019/elec/elec1601.html.