

Introduction to Machine Learning and Data Mining

COMP5318/COMP4318 Machine Learning and Data Mining

semester 1, 2026, week 1

Lecturer: Nguyen Tran

Some slides prepared by: Irena Koprinska

Reference: Witten ch.1, Tan ch. 1





Unit Coordinator	Lecturer	Lecturer
Nguyen Tran Email: nguyen.tran@sydney.edu.au	Michael Podbury Email: michael.podbury@sydney.edu.au	Tilda Yu Email: xiaotong.yu@sydney.edu.au

- Lectures
 - 2 hours weekly, 18:00-20:00 on Monday
 - There are always gaps in our understanding
 - Encourage discussion
- Tutorials (lecture content + code examples)
 - 1 hour weekly
 - Bring your laptop
 - **Allow to attend multiple sessions**
 - **Ask questions until you satisfy**

Important contents will be emphasized 3 times:

Lecture

- Tutorial
- Following lecture

Be critical of ideas, not people.

- The main place for this course is the Canvas website; we will use it for:
 - all teaching materials (unit outline, lecture slides, tutorial notes, tutorial solutions, assignments)
 - posting marks
- We will also use the discussion board Ed Discussion which is linked to Canvas

Type	Description	Weight	Due	Length	<u>Use of AI</u>
Written exam 🏹	Final exam Exam	60%	Formal exam period	2 hours	AI prohibited
Outcomes assessed: LO1 LO4					
Out-of-class quiz	Weekly homeworks Homework exercises	5%	Multiple weeks	-	AI allowed
Outcomes assessed: LO1 LO4					
Practical skill 🌸	Assignment 1 Computer program and report. Group assignment.	15%	Week 07 Due date: 17 Apr 2026 at 23:59	-	AI allowed
Outcomes assessed: LO1 LO2 LO3 LO4					
Creative work 🌸	Assignment 2 Computer program and report. Group assignment.	20%	Week 11 Due date: 15 May 2026 at 23:59	-	AI allowed
Outcomes assessed: LO1 LO2 LO3 LO4					
🏹 = hurdle task ? 🌸 = group assignment ?					

- The Quiz submission box closes at every Tuesday 5pm
- Multiple attempts are allowed
- Late and missed submissions
 - No late submissions are allowed, no extensions are possible
 - The only way to make up for a missed homework is to submit a special consideration via the University Special Consideration System
 - If it is approved (you need to provide a medical certificate or other valid evidence), the decision will be “Mark adjustment” – the mark for the missed homework will be set to the average of the other homeworks

- Two assignments
 - Programming assignments using Python and its machine learning libraries
 - Given a problem, you need to apply machine learning algorithms to solve it
- Assignment 1 (15%)
 - groups of 3 students
 - Computer program + short report or answering a few questions
 - Submitted in Canvas
- Assignment 2 (20%)
 - groups of 4-5 students
 - Computer program and report
 - Submitted in Canvas

- Check on Canvas for Assignment due dates.
- Late submissions – allowed up to 3 days late
 - Late penalty of 5% per day will apply
 - Assignments submitted more than 3 days late will not be accepted
- **Important:** Running the program also takes time. Start working on the assignments early.

- Exam: 60% (individual), during the examination period
 - Supervised exam
 - In-person, paper-based
 - Duration: 2 hours
- A minimum of 40% on the exam is required to pass the course – School of Computer Science policy.

- You need good programming skills for this course
- We expect that all students have a background in at least one programming language, preferably Python
- We will use Python and its machine learning libraries, e.g. sklearn
- If you don't know Python or haven't used it recently, you need to catch up
 - There are many Python books and resources available on the web that you can use
 - We have prepared a short Python refresher document – see Canvas

- We will use Jupyter Notebook
 - A web-based interactive programming environment
 - See the document on Canvas on how to install it on your computer
- In summary - 2 important resources on Canvas related to the practical part of this course:
 - How to install Jupyter Notebook
 - Python refresher (short document)

- During the tutorials we will do **practical exercises** using Python and its machine learning and neural network libraries
 - The exercises will be provided in a Jupyter Notebook format (.ipynb)
 - They are very detailed, we hope you will find them useful
 - Sometimes it may not be possible to finish all exercises during the 1-hour tutorial - you should do this at your own time
- For some of the weeks, we will also have **theoretical exercises** - paper-based exercises and calculations, testing your understanding of the algorithms
 - We will do some of them at the lectures, the rest should be done at your own time
 - Make sure that you do all theoretical exercises as they are similar in style to the exam questions
- The solutions for both types of exercises will be provided after the last tutorial (Friday evening)

- **Textbooks:**

- Pang-Ning Tan, Michael Steinbach, Anuj Karpathe and Vipin Kumar (2019). *Introduction to Data Mining*, 2nd edition. Pearson. (You can also use the previous edition)
- Ian H. Witten, Eibe Frank, Mark Hall and Christopher J. Pal
- *Data Mining - Practical Machine Learning Tools and Techniques*, 4th edition, Morgan Kaufmann, 2017 (You can also use the 3rd edition)

- **Books for the practical part using Python:**

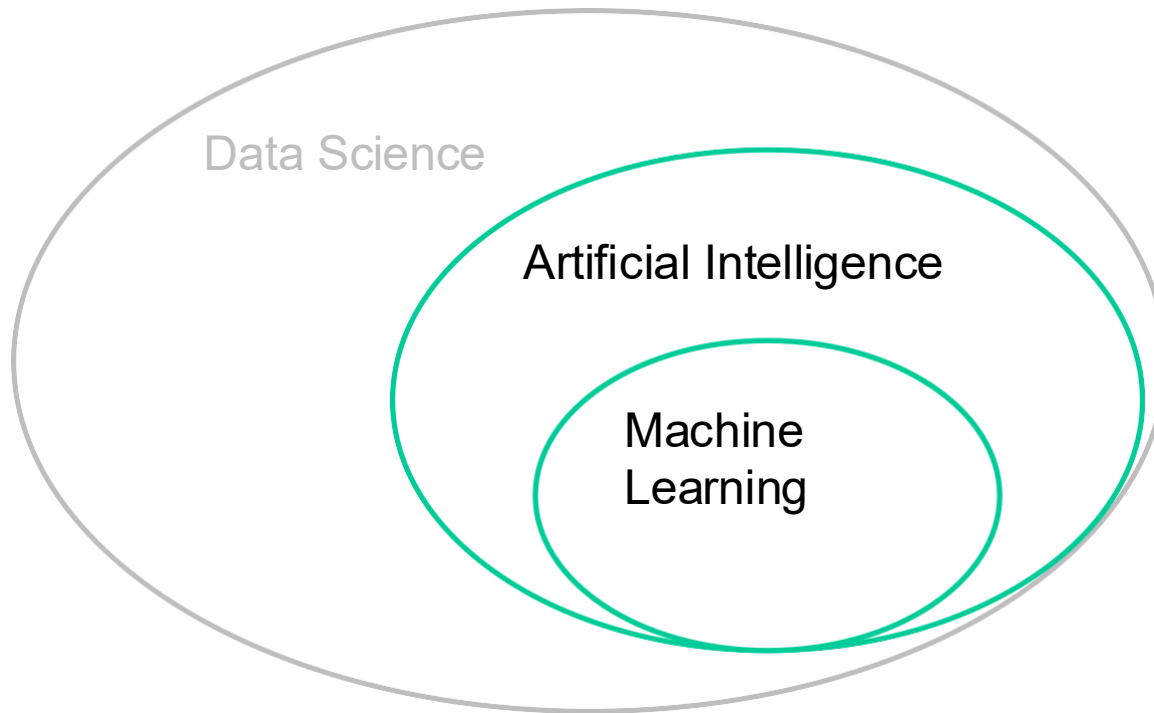
- Andreas C. Mueller and Sarah Guido (2016). *Introduction to Machine Learning with Python: a Guide for Data Scientists*, O'Reilly.
- Aurelien Geron (2022). *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow*, 2nd edition, O'Reilly. (You can also use the previous edition)
- All books are available from the library, both as hard copies and online versions, except Tan which is available as a hard copy only

Special Considerations (SC) due to Illness or Misadventure

- There is a centralized University system:
<http://sydney.edu.au/special-consideration>
- Applications are submitted online, after login to “myUni”
- You are required to submit the SC form within **3 working days** from the date when the assessment was due
- Applications are assessed by the University Student Administration Services (SAS) unit



Introduction to Machine Learning



Machine Learning vs Artificial Intelligence

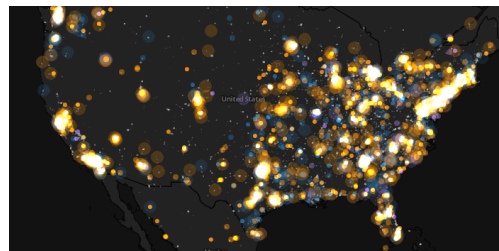


Data is everywhere

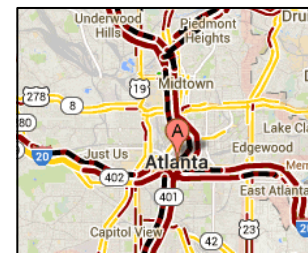
- Society produces and stores **huge amounts of data**
 - Due to automated data collection tools and sensors, mature database technology, cheaper and more powerful computers
 - Sources: business, science, medicine, economics, environment, web, etc.
- Examples:
 - purchase data – supermarket, department stores, online stores – e.g. Amazon handles millions of visits a day
 - bank/credit card usage data
 - web data – Google, Facebook; other social networking sites
 - telephone call details, government statistics, traffic data



E-Commerce



Social Networking: Twitter



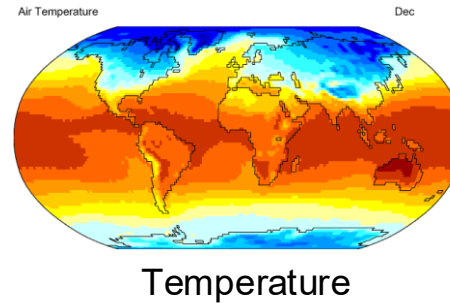
Traffic

amazon.com

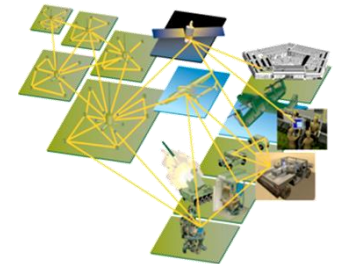
Google

facebook

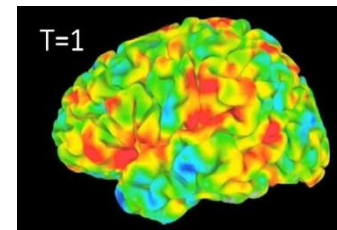
- Scientific data:
 - telescopes scanning the skies
 - remote sensors on satellites
 - weather data



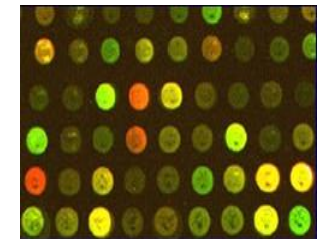
Sky survey data



Sensor networks



fMRI brain data



Gene expression data

- medical records and scans
- biological data (high-throughput) – cytometry, gene expression

- Data is collection of examples (also called *instances, records, observations, objects*)
- Examples** are described with attributes (*features, variables*)

Examples

<i>Tid</i>	<i>Refund</i>	<i>Marital Status</i>	<i>Taxable Income</i>
1	Yes	Single	125K
2	No	Married	100K
3	No	Single	70K
4	Yes	Married	120K
5	No	Divorced	95K
6	No	Married	60K
7	Yes	Divorced	220K
8	No	Single	85K
9	No	Married	75K
10	No	Single	90K



Introduction to Data Mining

Data mining is the process of extracting and finding patterns in massive data sets (involving methods at the intersection of machine learning, statistics, and database systems).

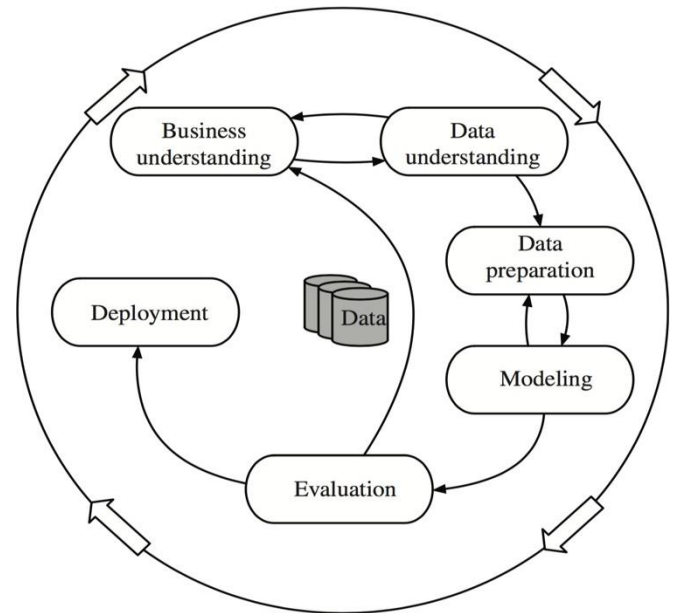


**What steps are usually involved in
the data mining process?**

Cross-industry standard process for data mining (CRISP-DM) is an open standard process model that describes common approaches used by data mining experts

Takeway: CRISP-DM breaks the process of data mining into six major phases:

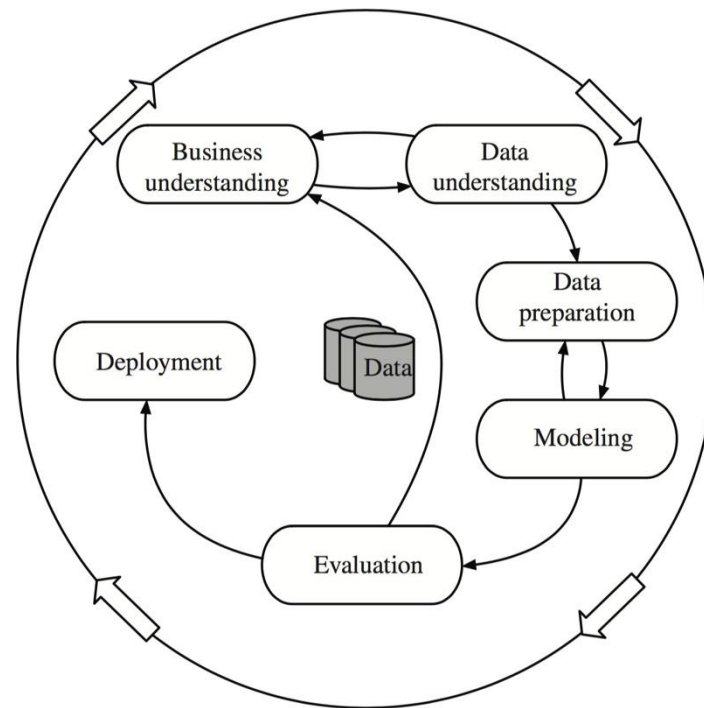
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment



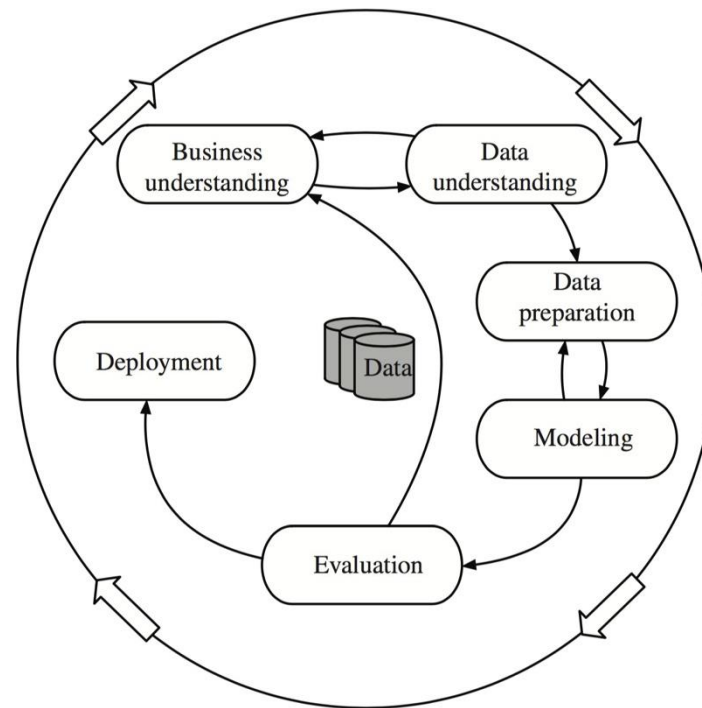
The sequence of the phases is not strict and moving back and forth between different phases is usually required.

1) Business understanding

- Investigating the business objectives and **requirements**
- Determining **what kind of data** should be collected



- Get an **initial dataset**; is it suitable for further processing?
- If the data quality is poor, **collect more data** based on more strict criteria
- Gain insights from data and **review the objective** – can useful information be extracted?



- Two common types of attributes:

- nominal (categorical)* - their values belong to a pre-specified, finite set of possibilities
- numeric (continuous)* - their values are numbers

outlook	temp.	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
?	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	?	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	?	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

nominal

sepal length	sepal width	petal length	petal width	iris type
5.1	3.5	1.4	0.2	iris setosa
4.9	3.0	1.4	0.2	iris setosa
4.7	3.2	1.3	0.2	iris setosa
...				
6.4	3.2	4.5	1.5	iris versicolor
6.9	3.1	4.9	1.5	iris versicolor
5.5	2.3	4.0	1.3	iris versicolor
6.5	2.8	4.6	1.5	iris versicolor
6.3	3.3	6.0	2.5	iris virginica
5.8	2.7	5.1	1.9	iris virginica
...				

numeric

3) Data preparation - preprocessing the data.

- Cleaning: data in real world is:
 - Incomplete, e.g. missing values
 - Noisy, e.g. containing errors or outliers
 - Inconsistent, e.g. in names

If there are missing values,
how should we handle them?

How do we know if there is noise?

outlook	temp.	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
?	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	?	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	?	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Takeaway: In general, cleaning data difficult and involves trade-offs.

3) Data preparation - preprocessing the data.

- Cleaning: data in real world is:
 - Incomplete, e.g. missing values
 - Noisy, e.g. containing errors or outliers
 - Inconsistent, e.g. in names
- Transformation: convert to common format; transform to new format; perform normalization, dimensionality reduction and feature selection

4) Modelling – **building intelligent methods (e.g., AI, ML methods) to extract knowledge (useful patterns) from it**

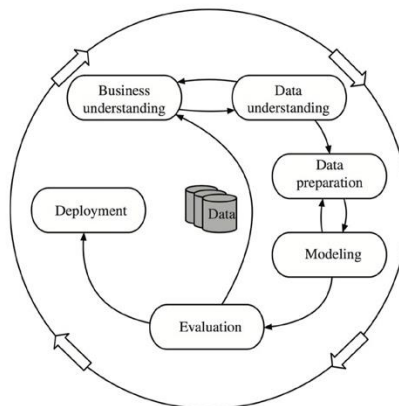
- Based on provided data, we will have three main types of machine learning methods:
 - Supervised learning - classification and regression
 - Unsupervised learning – clustering
 - Reinforcement Learning
- We will cover algorithms for supervised and unsupervised learning, and also for reinforcement learning

5) Evaluation

- How **good is the performance**? E.g. accuracy.
- Are the **patterns meaningful and useful**, or just reflecting spurious regularities?
- If the performance is poor, reconsider the project and return to step 1)
- If the performance is good -> deploy it in practice

6) Deployment

- Typically requires **integration into a larger software system** by software engineers
- May be necessary to re-implement the model in a different programming language



AvgTransactionAmount	TransactionsPerDay	GeoDispersion
87	4	1
78	3	2
90	3	2
72	4	1
82	4	2
60	3	2
83	1	2
2737	8	2
2227	9	2
2921	10	2
2079	8	2
2676	9	1
3157	9	1
...	...	...
...	...	...
2642	9	2
2143	9	1
2697	7	2
341	5	9
477	6	10
601	7	8
673	4	11
260	7	8
433	3	8
405	5	10
369	6	10
853	5	10
581	4	10
248	5	9
684	5	8

Bank transaction data:

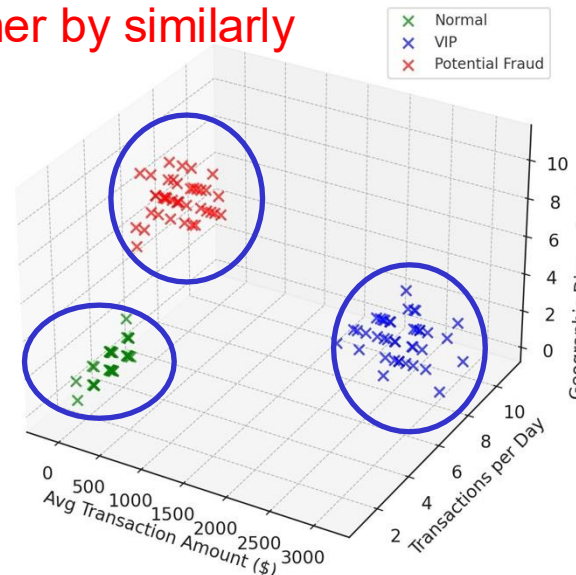
- Average transaction amount
- Transactions per day
- Geographic dispersion (number of different countries transactions occur in)

With such a dataset, what can we do?

Can we detect fraud?

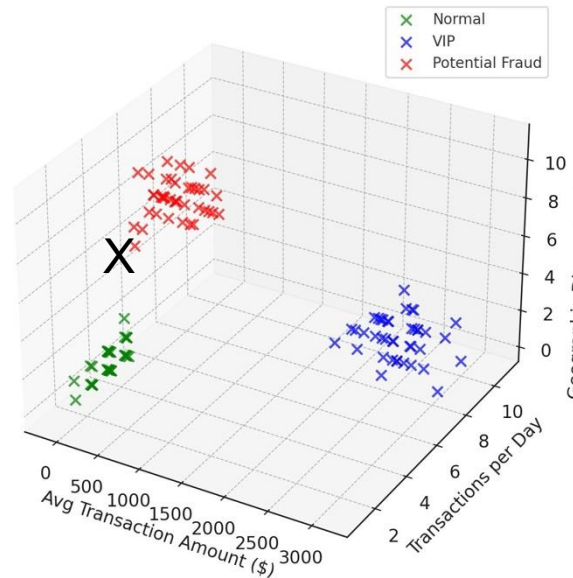
AvgTransactionAmount	TransactionsPerDay	GeoDispersion
87	4	1
78	3	2
90	3	2
72	4	1
82	4	2
60	3	2
83	1	2
2737	8	2
2227	9	2
2921	10	2
2079	8	2
2676	9	1
3157	9	1
...	...	...
...	...	...
2642	9	2
2143	9	1
2697	7	2
341	5	9
477	6	10
601	7	8
673	4	11
260	7	8
433	3	8
405	5	10
369	6	10
853	5	10
581	4	10
248	5	9
684	5	8

Takeaway: Clustering is to group data together by similarity



- **Cluster 1:** Normal customers (consistent spending patterns)
- **Cluster 2:** High spend, local usage → *VIP customers*
- **Cluster 3:** Low spend, suddenly many countries → *Maybe fraud*

- Ex.1: Targeted marketing
 - Segment customers into groups with distinct characteristics
- Ex. 2: Gene clustering
 - Find genes with similar structure
- Ex. 3: Document clustering
 - Find groups of documents that are similar to each other based on their content



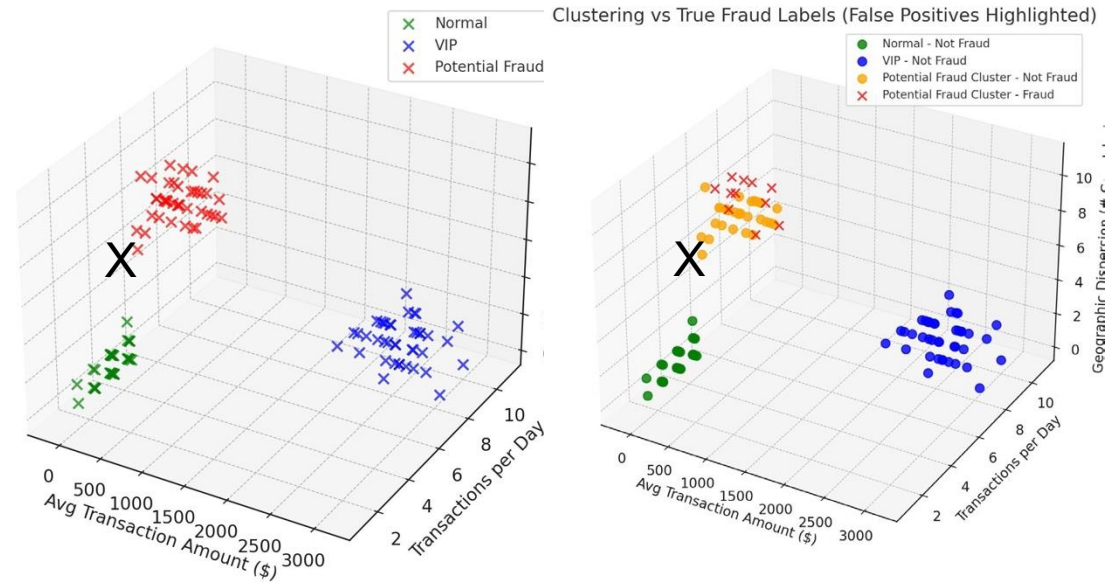
- **Cluster 3:** Low spend, suddenly many countries → *Maybe fraud*

Is the new record X a fraud?



Unsupervised learning - clustering

AvgTransactionAmount	TransactionsPerDay	GeoDispersion	IsFraud
87	4	1	0
78	3	2	0
90	3	2	0
72	4	1	0
82	4	2	0
60	3	2	0
83	1	2	0
2737	8	2	0
2227	9	2	0
2921	10	2	0
2079	8	2	0
2676	9	1	0
3157	9	1	0
...	...	...	...
...	...	...	...
2642	9	2	0
2143	9	1	0
2697	7	2	0
341	5	9	0
477	6	10	0
601	7	8	0
673	4	11	1
260	7	8	0
433	3	8	0
405	5	10	1
369	3	10	1
853	5	10	1
581	4	10	1
248	5	9	0
684	5	8	0



Takeaway: Be aware of the gap between the data provided (which has no labels about fraud) and the objective (predicting fraud) we want to achieve.

Attributes (features)

Labels (Targets)

AvgTransactionAmount	TransactionsPerDay	GeoDis person	IsFraud
87	4	1	0
78	3	2	0
90	3	2	0
72	4	1	0
82	4	2	0
60	3	2	0
83	1	2	0
3157	9	1	0
...	...	...	...
...	...	...	...
2642	9	2	0
2143	9	1	0
2697	7	2	0
341	5	9	0
477	6	10	0
601	7	8	0
673	4	11	1
260	7	8	0
433	3	8	0
405	5	10	1
369	3	10	1
853	5	10	1
581	4	10	1
248	5	9	0
684	5	8	0

When is it fraud?

Relationships: isFraud=1 when
 $\text{GeoDispersion} \geq 10$ and
 $\text{TransactionsPerDay} \leq 5$.

Takeaway: Supervised learning is used to automatically discovers relationships (useful patterns) from attributes to labels.

When labels are discrete (e.g., 0, 1, 2), supervised learning is called a **classification** method.

When labels are continuous (e.g., real-valued numbers), supervised learning is called a **regression** method.

- Given: a set of (labelled) examples $\{x,y\}$
 - x – input vector, y - target output
- Task: learn a function (model) that maps $x \rightarrow y$ and can be used predictively
 - to predict the value of y given the values of x for new, unseen examples
- Why is it called supervised?
- Two types of supervised learning
 - **Classification deals with discrete outcomes** (eg, yes/no, categories), while
 - **regression** handles continuous values (eg, price, temperature).

Example 1: Email Spam Detection

Task: Predict whether an email is spam or not spam.

Given a Dataset:

- **Features:** Email content (words, frequency, sender address)
- **Labels:** {Spam, Not Spam} → classification problem

Example 2: House Price Prediction

Task: Predict the selling price of a house.

Given a Dataset:

- **Features:** Number of bedrooms, location, floor area, year built
- **Targets:** Continuous values (price in dollars) → regression problem

Example 3: Handwritten Digit Recognition

Task: Identify the digit (0–9) from an image.

Given a Dataset:

- **Features:** Pixel values from the digit image
- **Labels:** {0, 1, 2, ..., 9} → classification problem

- Predict the electricity demand
 - Data: previous electricity demand, weather data and weather forecast data for the future days
 - Important to prevent blackouts and ensure reliable supply of electricity; also important for the economical and efficient operation of the electricity grid and for supporting the electricity market participants
 - Short- and long-term predictions - for the next few hours, next day, next week etc.; every 5 min, 30 min, 60 min, etc.
- Predict the exchange rate of AUD
 - Data from previous days, economical indicators, political events
- Predict retirement savings
 - Data: current savings and market indicators
- Predict the house prices in Sydney in 2030
- Predict the stock market index
- Predict wind velocity based on temperature, humidity, pressure

Week	Lecture
1	Administrative matters and course overview. Introduction to machine learning and data mining.
2	Nearest neighbour. Rule-based algorithms.
3	Linear regression. Logistic regression. Overfitting and regularization.
4	Naive Bayes. Evaluating machine learning methods.
5	Decision trees. Ensembles.
6	Support vector machines. Kernels. Dimensionality reduction.
7	Neural networks - perceptrons and back-propagation algorithm.
8	Deep neural networks I: CNN, RNN, and Transformer
9	Deep neural networks II: CNN, RNN, and Transformer
10	Clustering
11	Markov models
12	Reinforcement learning
13	Revision and preparation for the exam.

Welcome and
information to support
your University journey



THE UNIVERSITY OF
SYDNEY

CRI005 00026A



Outline

- Staying safe in the event of an emergency
- Your role in ensuring a safe campus
- Accessing tools and support
- Understanding your academic integrity responsibilities
- Using generative AI
- Completing your mandatory Canvas modules

We recognise and pay respect to the Elders and communities – past, present, and emerging – of the lands that the University of Sydney's campuses stand on. For thousands of years they have shared and exchanged knowledges across innumerable generations for the benefit of all.



Emergency and evacuation

In the unlikely event of a serious emergency, such as a serious ill or injured person or threat, call **triple zero (000)** for emergency services.

During an emergency, you may be required to evacuate. All University staff, students and visitors are required to **respond to emergency alarms** and **follow instructions**.



Check for any sign of immediate danger



Shutdown equipment/processes that cannot be left unattended



Locate your closest exit



Follow the green exit signs to find your way out of the building (**do not use lifts**)



Proceed to the assembly area, a safe distance from the building, while waiting until the emergency is over.



More information on staying safe in the event of an emergency is available at:
www.sydney.edu.au/about-us/campuses/emergencies-and-personal-safety.html

Emergency lockdown or shelter in place

In some circumstances, we might be asked to remain inside the building for our own safety. We call this a lockdown or shelter-in-place.



You will be notified of an emergency lockdown via **building announcements or other mechanisms** including **verbally by Protective Services or Building Emergency Wardens**.



It is important that you **follow directions from a Warden, Protective Services and/or Emergency Services**, remain calm and **stay inside the building**.



If your life is being threatened, or if someone is seriously ill or injured, or if you just witnessed a serious incident **dial triple zero (000) for emergency services** (fire/ambulance/police).



More information on staying safe in the event of an emergency is available at:
www.sydney.edu.au/about-us/campuses/emergencies-and-personal-safety.html

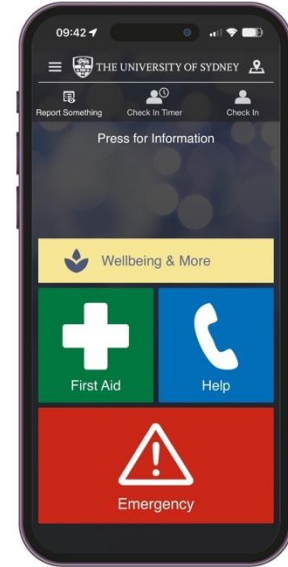
Safezone App

SafeZone is a **free safety app** for all University of Sydney students and staff. It connects you quickly with Protective Services, emergency teams, and sends real-time safety alerts.

Get started with SafeZone by **downloading the app** and logging in with your **university email**.

While on campus, you can access features including:

- **Emergency assistance:** Connect instantly with Protective Services
- **First aid:** Request medical help
- **Timed Check-In:** Set a timed check-in when working or studying alone
- **Call Security:** If you're locked out, lost on campus or need an escort
- **Report something:** If you notice something concerning or unusual
- **Wellbeing support:** Connect with student support services
- **Location sharing:** So help can reach you faster
- **Push notification alerts:** Keep informed about critical incidents



More information on SafeZone is available at:
sydney.edu.au/safezone

Your role in ensuring a safe campus

We are committed to providing a safe and healthy learning environment that is free from bullying, unlawful harassment and discrimination.



You can ensure the safety of yourself and your peers by understanding your rights and responsibilities, which are detailed in our Student Charter.



If you experience or witness any form of behaviour that breaches the Student Charter or our codes of conduct, you can report it to our Student Affairs team.



Our support services, including the Wellbeing team and Safer Communities Office, can support you in dealing with any threats to your safety. You can connect with these teams via the Student Portal.



If your class is full...

Once a class is physically full, the lecturer may ask students not to enter. This is to ensure the safety of everyone. For example, sitting or standing on the stairs or in walkways may block emergency exits.



You will have access to a recording of the lecture shortly after it finishes as well as online resources including the lecture notes.



You will not be penalised for non-attendance.



Your lecturer will let you know if there are additional classes or other options.

Accessing tools and support

Accessing tools and support for your studies and wellbeing

Stay Organised. Stay Connected. Stay Supported.

Make uni life easier with the Student Portal & Sydney Uni App



Stay organised – Use the new **Study Dashboard** to plan for your semester and view your assignments and to-do lists all in one place.



Stay supported – Need study help, wellbeing resources, or student services? The **Service and Support page** has everything you need



Stay connected – Find events on **What's On**, track your classes and exams with the **Calendar**, and find your way with **Campus Maps**.



Download the Sydney Uni app

Visit myuni.sydney.edu.au or download the [Sydney Uni app](#).

Accessing assistance at our in-person hubs

Need in person assistance on campus? Visit our Student Services Hubs, where we can help you with:

- Completing enrolment
- Understanding your timetable
- Course & Degree planning
- Candidature variations
- ICT support
- Peer and learning support
- General advice to thrive at The University of Sydney

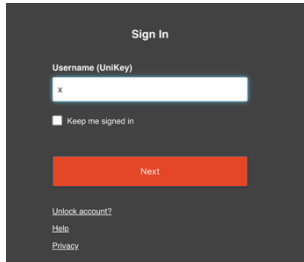
Student Services Hubs are open **Monday to Friday, 9 am to 5 pm** in **Fisher Library** , the **Belinda Hutchinson Building**  and the **Susan Wakil Health Building** .

You can collect your Student ID card from our Cards Hub, located in Jane Foss Russell **after the date in your 'Ready to Collect' email** . Cards cannot be printed prior to this date. For instructions on initiating a card request, visit the [Student Cards](#) website.



Managing access to our University systems

Multifactor Authentication (MFA) is an additional security step to verify your identity at the University.

A screenshot of a dark-themed 'Sign In' form. It features a 'Username (UniKey)' label above a white input field containing the letter 'x'. Below the input field is a checkbox labeled 'Keep me signed in'. A large orange 'Next' button is positioned below the checkbox. At the bottom, there are links for 'Unlock account?', 'Help', and 'Privacy'.

Step 1

*Sign in with your username
(UniKey and password)*



Step 2

*Verify your identity using
your smartphone*



*For more information
Scan the QR code*

Remember:

- Set up a secondary email, phone number and security question in your Okta profile so that you can reset your Unikey.
- Always have your Okta enabled device with you and do not uninstall the Okta verify app.
- If you change your mobile device, install Okta Verify on your new device before removing it from your old one.

Academic integrity and the use of AI

Understanding your academic integrity responsibilities

Academic integrity refers to behaving honestly, ethically and responsibly in relation to all elements of your study at the University, including assessments.

It's important that you:



Always submit your own work, sit your own tests, and take your own examinations.



Acknowledge any contributions in your assignment which are not your original thoughts, ideas or words.



Complete the Academic Honesty Education Module (AHM), which all new students must complete by census date (31 March 2026).



Listen to your coordinator's instructions about use of Generative AI in assessments



Be alert to individuals or companies that claim to offer students external academic assistance. Many are illegal 'contract cheating' services.



Scan to enrol in AHM and view
Academic Integrity resources

Using generative AI

Generative AI tools, such as Microsoft Copilot Chat, can create digital content, including text, images, and video. Different guidelines for use of generative AI apply to learning and assessment:



Generative AI for general learning

- You can use generative AI to help you learn. For example, it can explain things, apply knowledge, plan your study and make practice questions. Use our student co-designed AI resource so you are aware of its strengths and limitations.



Generative AI for assessment

- Check your unit outline for whether you can use generative AI tools in your assessment:
 - **Secure assessments** (in-person, supervised): generally no AI allowed unless express permission.
 - **Open assessments** (unsupervised): generally AI is allowed.
 - Failing to acknowledge AI use is an academic integrity breach.
- When using generative AI tools, don't provide content generated by other students, others' intellectual property (unless with permission), or personal or health information. Your unit coordinator can provide guidance on this.



AI resource made by students:
<https://bit.ly/students-ai>

Which generative AI tools are best to use?

The University provides **free** access to Microsoft Copilot Chat to accelerate your learning and help you in those assessments where use of generative AI is allowed.



Why use Microsoft Copilot Chat?

- It is a University-supported, industry standard tool powered by up-to-date AI models
- It keeps your data secure and protects your privacy
- Log in with your UniKey using Okta so that you are using Copilot in protected mode, which introduces necessary guardrails to protect you within the system



Why not use other tools like ChatGPT and DeepSeek?

- Watch out for your privacy and the security of your personal and other data if using a free or emerging tool like ChatGPT and DeepSeek



Powerful and secure
<https://copilot.microsoft.com/>

Mandatory Canvas modules

Completing your mandatory Canvas modules

We've developed a series of Canvas modules to help you understand what to expect from the University and what is expected from you. They include:

- *Respect@Sydney 2026*, explores key concepts related to gender-based violence, consent, respectful relationships, and support and reporting options (mandatory for commencing students, continuing students will be provided more information early in semester 1, 2026 on their requirements for modules).
- *Academic Honesty Education Module* (mandatory for commencing students).
- *Engaging with the Student Charter*, a module where you can unpack the values and expectations outlined in our Student Charter, including information about penalties for breaching our codes of conduct and displaying threatening, discriminatory or racist behaviour.
- *Anti-Slavery Awareness Module*.

Many of these modules are mandatory for commencing students and your results may be withheld if you do not complete them.

Useful links

- www.sydney.edu.au/students/new-students.html
- www.sydney.edu.au/students/glossary.html
- www.sydney.edu.au/students/support.html
- www.sydney.edu.au/students/academic-integrity.html
- www.sydney.edu.au/students/exams.html
- www.sydney.edu.au/students/scams.html
- www.library.sydney.edu.au/visit
- www.sydney.edu.au/students/log-in-to-university-systems.html
- www.sydney.edu.au/students/student-it.html
- www.sydney.edu.au/students/key-dates.html

Academic Integrity: What you need to know



THE UNIVERSITY OF
SYDNEY

CRICOS 00026A TEQSA PRV12057



Academic integrity is...

*“a commitment to five fundamental values: **honesty, trust, responsibility, fairness, respect** ... plus the **courage** to act on them even in the face of adversity...”*

The Fundamental Values of Academic Integrity, [International Center for Academic Integrity \(Third Edition\)](#)

Acting with *academic integrity* means:

- Following ethical principles in all dealings at university
- Following the rules set for assignments and exams
- Doing your own work and never 'outsourcing'
- Acknowledging *any* contributions which are not your original thoughts, ideas or words
- When working in groups, doing so legitimately – i.e., not colluding

Academic integrity goes to the heart of what a university's role is: educating students, advancing research, contributing to the world's knowledge, and qualifying people to do particular jobs. Without integrity, a university's activities lose meaning and value.

Types of academic integrity breaches

Plagiarism

Presenting work that is not your own without acknowledging the original source.

Fabricating information

Making up information, such as experimental or interview data. It can also include inventing sources by citing publications that are incorrect or that don't exist.

Contract cheating

This involves having someone else complete all or part of your work for you. It includes

- buying a completed assignment
- having someone else write part or all of an assignment
- using model answers from a tutoring company
- engaging in the buying, selling, and swapping of assignments.

Misuse of AI

Using generative AI tools when they have been prohibited for assessment, or, where AI use is allowed, failing to acknowledge what tools you have used and how.

Collusion

Engaging in illegitimate cooperation with one or more students in the completion of assessable work.

Recycling

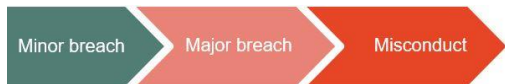
Re-submitting work that has already been assessed without your teacher's permission and for which you have already been given feedback.

Exam cheating

Breaching exam conditions, including using unpermitted materials or devices, communicating with others or allowing someone else to take the exam on your behalf.

The University's approach

The University **investigates** any student suspected of breaching the Academic Integrity Policy. Teachers monitor submissions for potential integrity issues and report any suspected breaches.



Substantiated breaches are classified in three categories:

minor breach, major breach, or misconduct.

The investigation process and penalties differ depending on the type of breach.

Potential penalties for breaches include:

- Marks deducted for an assignment or exam
- Requirement to do a 'development' activity
- Failure of the assessment task or the unit of study
- Suspension or exclusion (for more serious breaches)
- If you are an international student, you could even lose your student visa.

NOTE: The personal impact of investigations and adverse findings can go beyond the academic penalties.

Complete the *Academic Honesty Education Module* (AHM)

The Academic Honesty Education Module ('AHM') is compulsory for all new students, while other students are strongly encouraged to complete it.

The AHM provides essential information about academic integrity, and it typically takes about 1 hour to complete. It covers the key principles of academic integrity, the risks of engaging in breaches, and provides guidance on where to seek support.

New students must complete the AHM before the census date. The Module will appear in your Canvas dashboard at the start of semester.

Failure to complete the module could result in your academic results being withheld.

Be aware of contract cheating providers: *Know the risks*

Personal risks	
Academic consequences	Poor quality
Data security	Blackmail

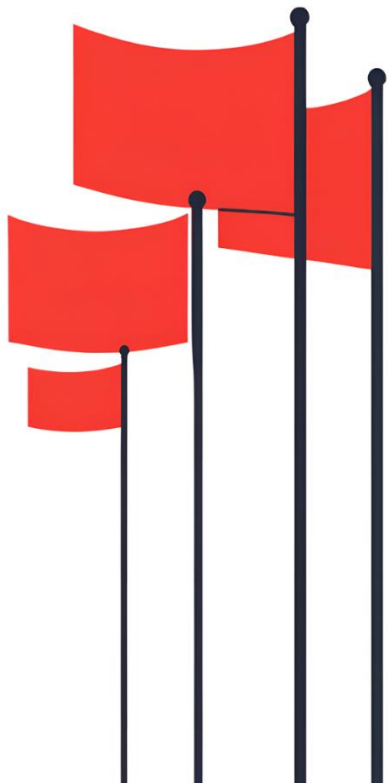
Contract cheating providers target students online and in person, sometimes even on campus such as at Welcome events. They often call themselves ‘tutoring’ and may offer other services alongside academic services such as housing advice, course advice, and appeals assistance.

Advertising or providing contract cheating services is against the law. Penalties can include fines of up to \$100,000 or imprisonment.

There are many risks....

- Contract cheating is considered **academic misconduct** and results in severe penalties – including failing the assignment or unit, suspension or exclusion.
- The quality of work is often **poor, badly written or wrong**. In many cases it doesn’t even get the student the desired grade.
- When you engage with these companies, you not only put yourself at risk of being caught, but also expose yourself to a range of other risks including **being blackmailed** by the company, harassed, and having your personal data/privacy compromised.

Signs you might be dealing with a *contract cheating service*



- You're approached or solicited to use the service in person, or online
- You're asked to pay for the service
- You're asked to share your Unikey or other personal details
- Advertisements for the service mention a specific USYD unit of study
- The service provides a 'guarantee' that you won't get caught
- The service is not acting in an open, transparent way
- You're added to a private group chat by the service
- You're given step-by-step instructions for how to complete an assessment
- The service is aggressive or threatening
- You feel that by using the service you are getting an unfair advantage

To report a service or if you find yourself in a difficult situation, contact the [Office of Educational Integrity](#)

QR code: How to spot a contract cheating service



Using generative AI

Generative AI tools, such as **Microsoft Copilot Chat**, can create digital content, including text, images, and video. Different guidelines for use of generative AI apply to learning and assessment:



Generative AI for general learning

- You can use generative AI to help you learn. For example, it can explain things, apply knowledge, plan your study and make practice questions. Use our student co-designed AI resource so you are aware of its strengths and limitations.



Generative AI for assessment

- Check your unit outline for whether you can use generative AI tools in your assessment:
 - For in person supervised tasks like exams – generally you will not be allowed to use AI unless your coordinator expressly permits it.
 - For other tasks, generally you will be able to use it, but you must appropriately acknowledge all use.
- When using generative AI tools, don't provide content generated by other students, others' intellectual property (unless with permission), or personal or health information. Your unit coordinator can provide guidance on this. Failing to acknowledge AI use is an academic integrity breach.



AI resource made by students:
<https://bit.ly/students-ai>

Which generative AI tools are best to use?

The University provides **free** access to Microsoft Copilot Chat to accelerate your learning and help you in those assessments where use of generative AI is allowed.



Why use Microsoft Copilot Chat?

- It is a University-supported, industry standard tool powered by up-to-date AI models
- It keeps your data secure and protects your privacy
- Log in with your UniKey using Okta so that you are using Copilot in protected mode, which introduces necessary guardrails to protect you within the system



Why not use other tools like ChatGPT and DeepSeek?

- Watch out for your privacy and the security of your personal and other data if using a free or emerging tool like ChatGPT and DeepSeek



Powerful and secure
<https://copilot.microsoft.com/>

File-sharing

File-sharing sites allow you to buy, trade or sell academic material. Using these sites puts you at risk of academic integrity breaches.

Do not upload:

- completed assessments (which can be used by other students to cheat) – this is a breach of the [Academic Integrity Policy](#).
- University course or teaching materials, including assignment questions – this is a breach of University copyright and the [Academic Integrity Policy](#).



Legitimate support

Contact your unit of study coordinator or teaching staff for guidance and support.

Learning Hub (Wentworth Building and online)

- [Learning Hub – Academic Language](#)
- [Learning Hub – Mathematics](#)
- [Learning Hub Canvas site](#) and [SCANA test](#)
- [Peer programs](#) – “Speak + Connect” and “Speak Up”

Library

- [Referencing guide](#)
- Meet with a [Peer Learning Advisor](#)
- [PeerConnect](#) (online chat)
- [Exam Ready](#) sessions

Other resources

- [AI in Education](#) (Canvas site on generative AI)
- [Study Skills page](#) of the student website
- [The Write Site](#) (Canvas site)
- [Power up your learning](#) (Canvas site)
- [Academic Progression](#)

Advice about integrity investigation

- [SRC Caseworkers](#) (undergraduates)
- [SUPRA Caseworkers](#) (postgraduates)

The Office of Educational Integrity
Education Portfolio
educational.integrity@sydney.edu.au

sydney.edu.au/students/academic-integrity

Questions?



THE UNIVERSITY OF
SYDNEY