VICTORIA UNIVERSITY OF WELLINGTON TE HERENGA WAKA

COMP309/AIML421 — ML Tools and Techniques

Week 1

Course Introduction

Dr Qi Chen

School of Engineering and Computer Science

Victoria University of Wellington

Qi.Chen@vuw.ac.nz

Meet the Team: Academics

Dr Qi Chen (Course Coordinator)

- Qi.Chen@vuw.ac.nz
- Week 1
  - Extensions and other admin

Dr. Junhong(Jennifer) Zhao
Postdoctoral Fellow
Week 1-3, and Week 5 Tutorials

Prof Ali Knott
Ali.Knott@vuw.ac.nz

- Week 2 – Week 6
  - Extensions and other admin
  - A1 – A3 and Project

A/Prof Marcus Frean
Marcus.Frean@ecs.vuw.ac.nz
Week 7 - Week 12
A3 & A4 and Project

Meet the Team: Tutors/Markers

- Have a team of Six tutors and Five markers

- Part of our tutor team members

Hengzhe Zhang

Amer Hussain

Fengyang Sun

- Will update later here https://ecs.wgtn.ac.nz/Courses/COMP309_2024T2/People

Meet the Team: Class Rep(s)?

- The key role of the Class Representative is to assist communication between staff and students in relation to course matters and to provide a point of contact for students.
- i.e. somebody to raise issues, comments, suggestions with us on your behalf!

- COMP 309: 1 rep
- AIML 421: 1 rep
- Email me please!

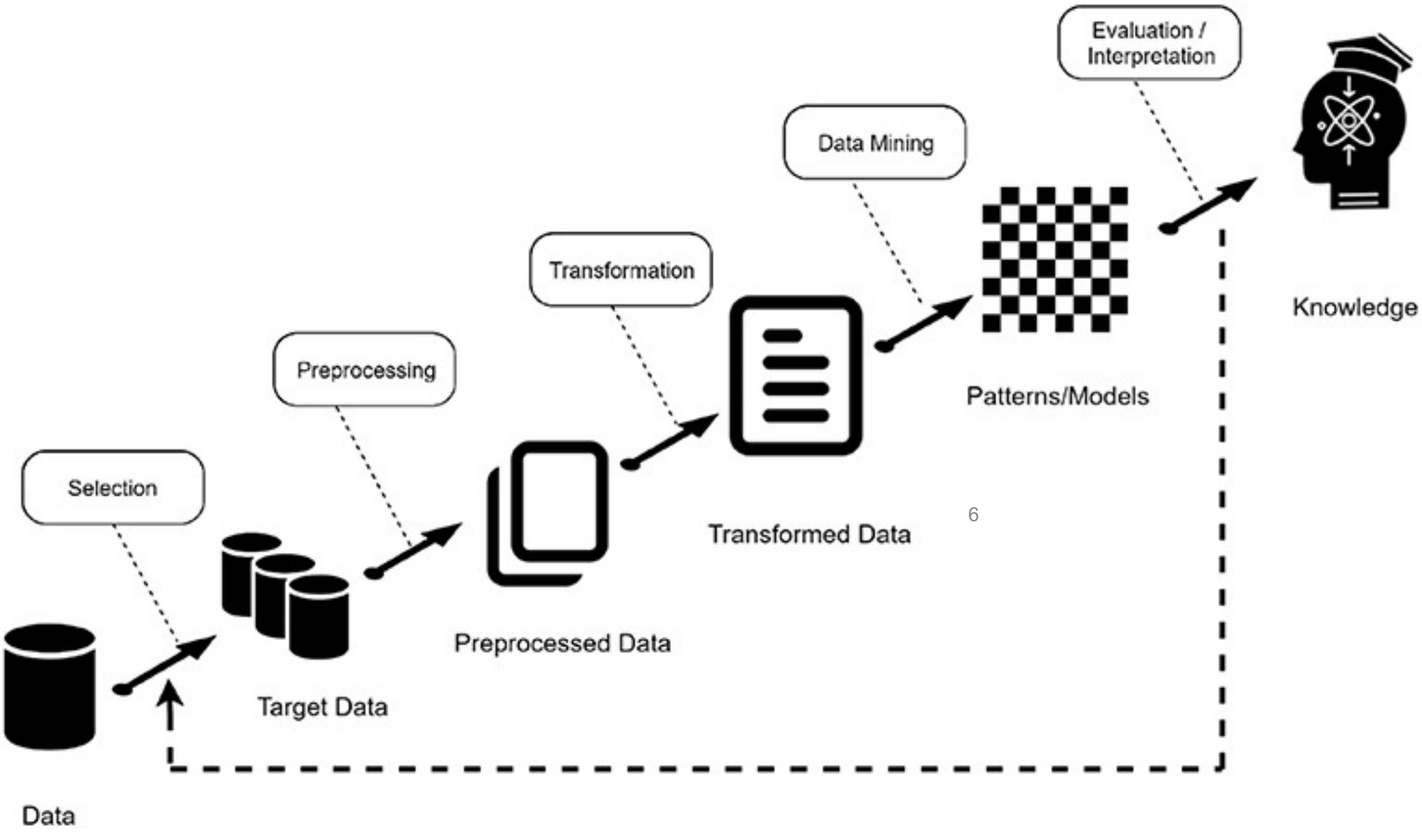
Volunteers?

Class Schedule

- Two lectures a week (Monday/Tuesday) 14:10 -15:00 pm, LT101, Maclaurin, Kelburn
  - Introduces crucial concepts and ideas!
- One tutorial a week (Thursday) 14:10-15:00 pm, MT228, Student Union, Kelburn
  - How to use tools/libraries to implement these ideas
- Helpdesks **every day** (from Week 2)
  - 12-1pm in CO242B
  - Not just for assignments
  - Extra 1-2pm helpdesks when assignment is due
- ALL learning materials are available online: https://ecs.wgtn.ac.nz/Courses/COMP309_2024T2/
- We *use* ***Nuku*** for announcement, class recording, and later course/teaching evaluation.

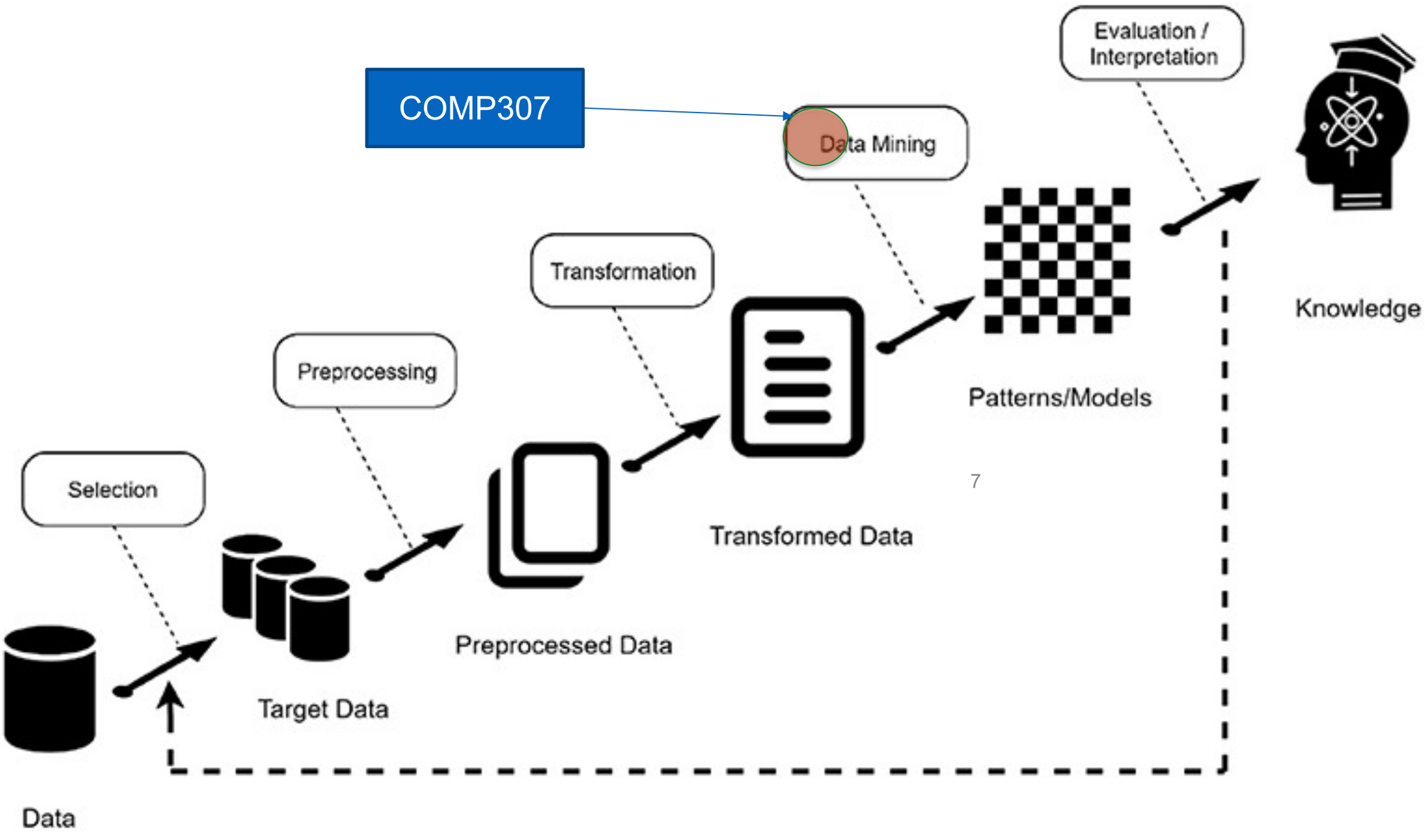
What is this Course?

- A journey through the whole data mining pipeline, with a focus on *machine learning*



How is this different from COMP307?

- Bigger picture – not focused on the *details* of AI algorithms



Assessment

- Four assignments
  - Two smaller ones (2 weeks): 15% each
  - Two larger ones (3-4 weeks): 20% each
- One final project
  - Due during assessment period (no exam): 30%
  - Deep learning + aspects from across the whole course
  - In-person marking

Assignment 1 (20%): due 2nd August 2024 (Friday Week 4) 23:59

Assignment 2 (15%): due 16th August 2024 (Friday Week 6) 23:59

Assignment 3 (20%): due 13rd September 2024 (Friday Week 8) 23:59

Assignment 4 (15%): due 27th September 2024 (Friday Week 10) 23:59

Project (30%): due 29th October 2024 (Tuesday Week 15) 23:59

- We will try to get your marks back within 2 weeks

Assessment-AIML421 Students

- Same number of assessment items
- No separate essay
- Each assignment and the project will have an additional question or two for AIML421 only
- In some cases, there may be a different (more complex) dataset etc.

Requirements

Mandatory Course Requirements

In addition to achieving an overall pass mark of at least 50%, students must:

- submit reasonable attempts for at least three of the four assignments, and
- submit a reasonable attempt at the final project.

Penalties

The penalty for assignments that are handed in late without prior arrangement (or use of "late days") is one grade reduction per day. Assignments that are more than one week late will not be marked.

Plagiarism

- Using somebody else's work as your own, without saying so.
- This includes anything not in the lectures/tutorials!

- **It's really easy to avoid.**
- Just **tell us** if you used a resource!! (and how much)
- It is never plagiarism if you are honest.
- The penalty for getting caught can be worse than getting no marks…it's not worth it.

Extensions/Deadlines

- We are using the “3 late days” model for this course applied to the four assignments but not the project.
  - Close deadlines
  - Computer problems
  - Getting really turnt at a party
  - …
- Thus, minor extensions will **not** be approved.
- Don’t *waste* your late days early on
- Medical/exceptional circumstances? Apply for extension via Submission System – preferably before the deadline

Workload

- 150 hours for 15 points
- 36 “contact hours”
- So 1% of assessment is roughly 1.14 hours

- Start the assessment early
  - You may not be able to do it all when it is released – but you can start!
  - Make the most of helpdesks – they are less busy early

- Make a calendar of ALL your deadlines & plan

Where to from here?

Week 1: Introduction and Overview

- Course Introduction:
- Overview of Machine Learning:
- Machine Learning Tasks:
- Tutorial: Tools: Jupyter, Python, Numpy, Pandas, Scikit-learn

Week 2: Models and Algorithms - Classification

- Classification and Classifiers:
- Tutorial: Scikit-learn for Classification

Week 3: Models and Algorithms - Regression, Dimensionality Reduction and Clustering

- Regression and Dimensionality Reduction:
- Clustering:
- Tutorial: Regression, Dimensionality Reduction, and Clustering

Week 4: Data Mining Process

- Data Mining:
- CRISP-DM:
- Tutorial:

Week 5: Exploratory Data Analysis (EDA)

- EDA:
- EDA Tools:
- Tutorial: EDA with Orange

Week 6: Data Pre-processing 1

- Data Preparation:
- Feature Manipulation-Feature Extraction and Feature Selection:
- Tutorial: Kaggle Competition:

Where to from here?

Week 7: Data Pre-processing 2

- Feature Engineering:
- Missing values:
- Tutorial: Feature Engineering , and Handling Missing Values

Week 8: Performance Evaluation

- Performance Metrics, part 1:
- Performance Metrics, part 2:
- Tutorial:

Week 9: Optimisation

- Optimisation:
- Gradient Descent:
- Tutorial: PyTorch and autograd

Week 10: Deep Learning(1)

- Neural Networks:
- Tutorial: MLP from scratch, using PyTorch

Week 11: Deep Learning(2)

- Deep Neural Networks, and images:
- Convolutional Neural Networks:
- Tutorial: CIFAR10 via a ConvNet

Week 12: A couple more things

- Ethics and hype:
- Generative models:

Now Let's Get On With It!