CIKM 2020
19-23 OCTOBER
GALWAY · IRELAND

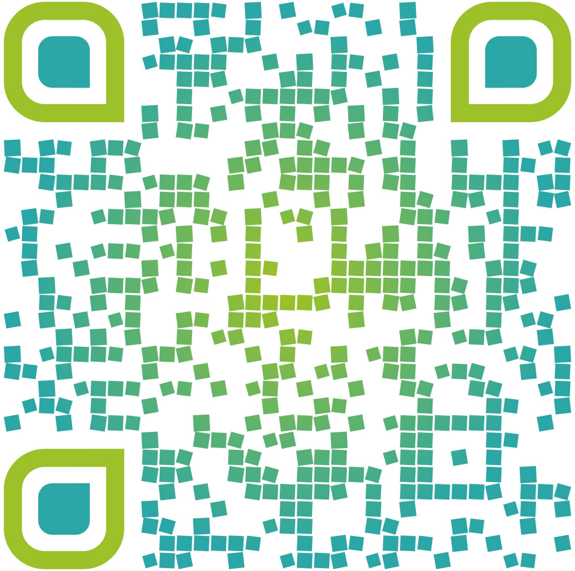
Challenges and Solutions to the Student Dropout Prediction Problem in Online Courses

Bardh Prenkaj, Giovanni Stilo and Lorenzo Madeddu

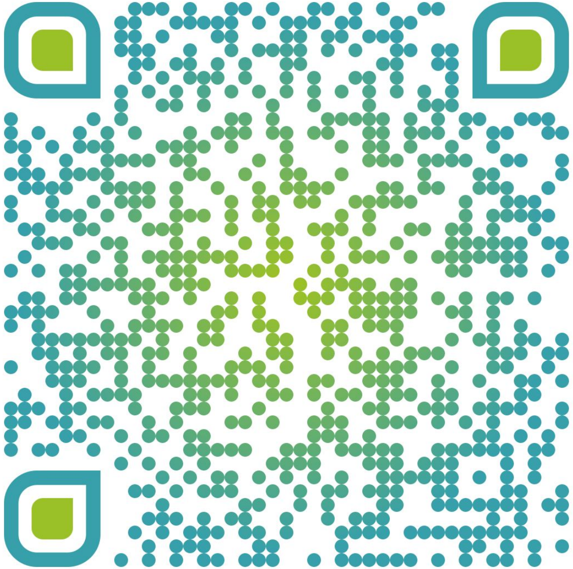
SAPIENZA
UNIVERSITÀ DI ROMA

Artificial Intelligence & Information Mining

Online Resources



Tutorials Website



Tutorials Slides

Part Zero
Tutors’ Presentation and Tutorial Roadmap

Tutors Biography

Bardh Prenkaj

- PhD in Machine Learning
- Winner of a co-founded scholarship in e-learning analytics
- First author of an exhaustive survey on SDP [1]
- Research interests:
  - Machine Learning
  - Temporal Data Mining
  - Time-series Analytics
  - Anomaly Detection

SAPIENZA
UNIVERSITÀ DI ROMA

Tutors Biography

Giovanni Stilo

- Associate Professor in Data Mining @ L'Aquila University
- Published works in the area of
  - Graph Mining and Network Medicine
  - Temporal Data Mining and Semantics-aware recommender systems
  - Anomaly Detection and Ensemble Mechanisms
  - Educational Data Mining and Learning Analytics
- Co-organiser of international workshops in top-tier conferences (ICDM, CIKM, ECIR)
- Editor and reviewer in TITS, TKDE, DMKD, AI, KAIS, AIIM
- Recently, formalised definitions in an exhaustive survey on SDP [1]

Tutors Biography

Lorenzo Madeddu

- PhD in Graph Mining and Network Biology
- Contributed as graph mining expert.
- Research interests:
  - Machine Learning
  - Network Medicine
  - Network Pharmacology

SAPIENZA
UNIVERSITÀ DI ROMA

Roadmap

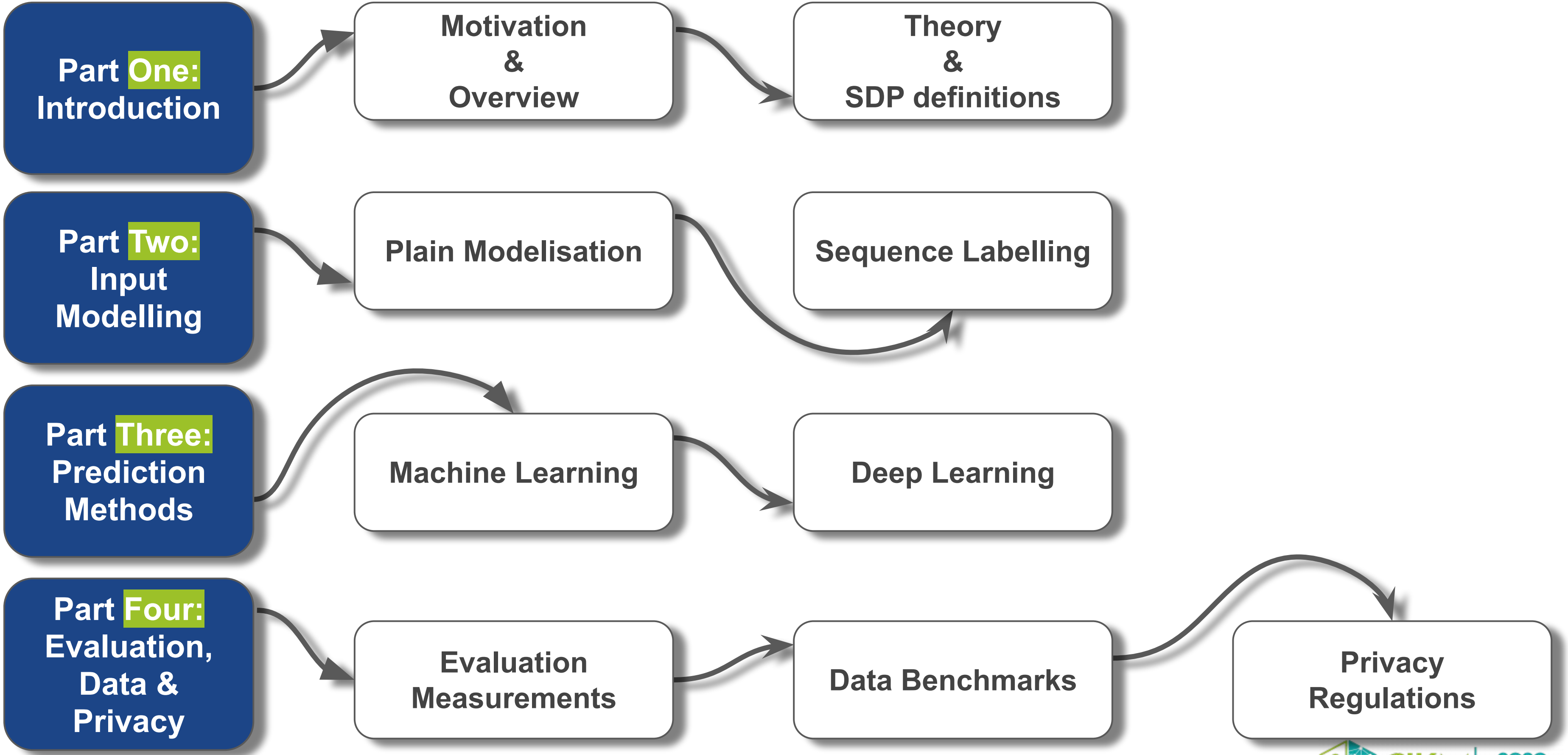
- Introduction and background theory [30 min]:
    - Motivation and overview [GS]
    - Theory and SDP definitions [GS]

- Input modelling techniques [50 min]
    - Plain Modelisation [LM]
    - Sequence Labelling [BP]

- Methods in SDP [60 min]
    - Machine Learning methods [LM]
    - Deep Learning methods [BP]

- Evaluation, datasets & privacy, open challenges [BP&GS] [40 min]

Roadmap



Part One: Introduction

Motivation & Overview

Theory & SDP definitions

Part Two: Input Modelling

Plain Modelisation

Sequence Labelling

Part Three: Prediction Methods

Machine Learning

Deep Learning

Part Four: Evaluation, Data & Privacy

Evaluation Measurements

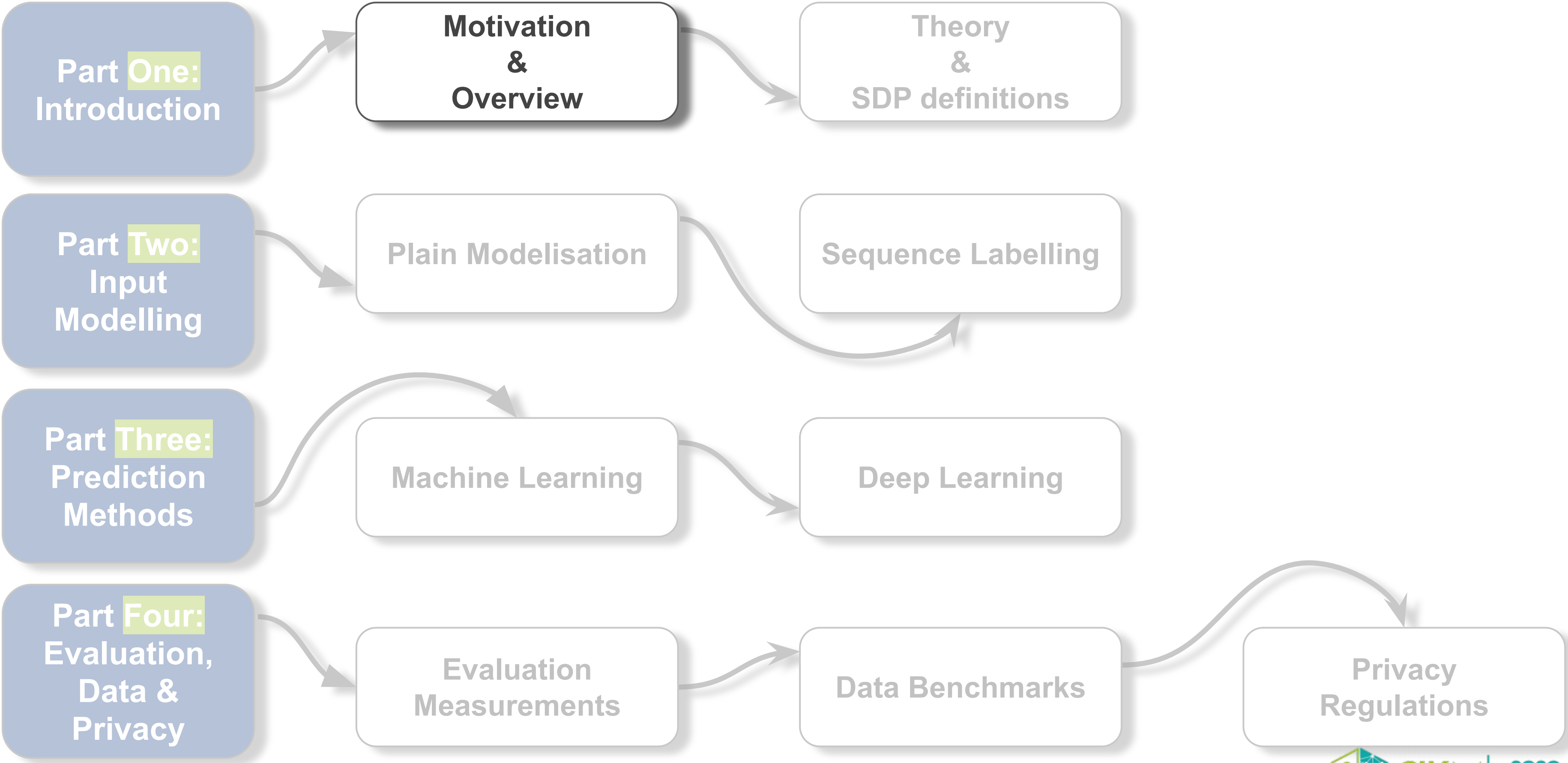
Data Benchmarks

Privacy Regulations

Part One

Introduction and Background Theory

Roadmap



Part One: Introduction

Motivation & Overview

Theory & SDP definitions

Part Two: Input Modelling

Plain Modelisation

Sequence Labelling

Part Three: Prediction Methods

Machine Learning

Deep Learning

Part Four: Evaluation, Data & Privacy

Evaluation Measurements

Data Benchmarks

Privacy Regulations

What is the SDP?

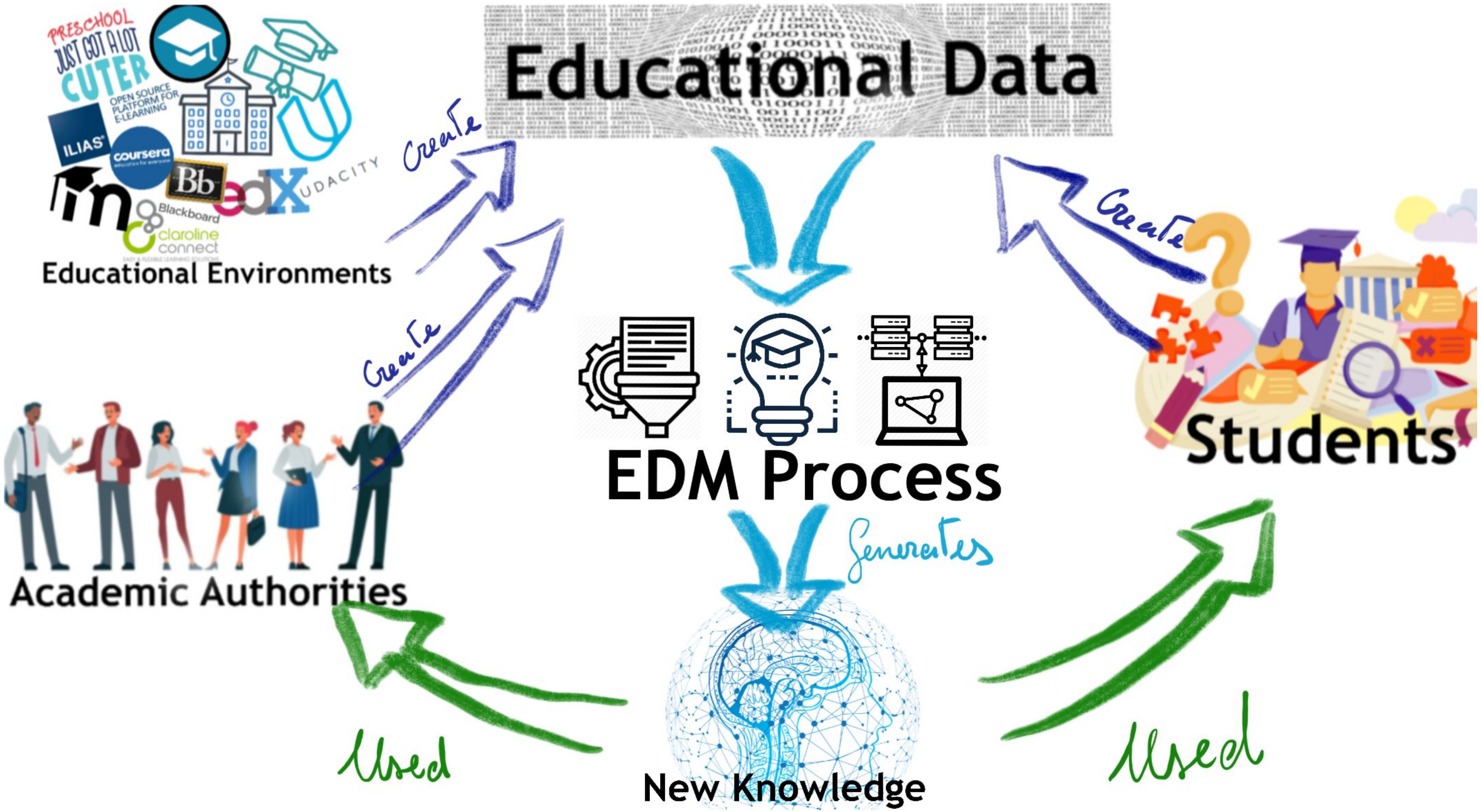
- The objective of SDP is to **analyse** student **dropout** in **distance** learning **environments**:
  - **modelling** student **behaviour** when interacting with **e-learning** platforms.

- Student dropout prediction (SDP) is a research topic in the **multidisciplinary** field of **Learning Analytics** (LA) [40].

- More precisely, it belongs to the area of **Educational Data Mining (EDM)** (see [5,27,60] for an overview of this field).

[5] Behdad Bakhshinategh, Osmar R Zaiane, Samira El Atia, and Donald Ipperciel. 2018. Educational data mining applications and tasks: A survey of the last 10 years. Education and Information Technologies 23, 1 (2018)
[27] S Hari Ganesh and A Joy Christy. 2015. Applications of educational data mining: a survey. In 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). IEEE.
[40] Usha Keshavamurthy and H. S. Guruprasad. 2014. Learning Analytics: A Survey. International Journal of Computer Trends and Technology (IJCTT) (2014).
[60] Alejandro Peña-Ayala. 2014. Educational data mining: A survey and a data mining-based analysis of recent works. Expert systems with applications 41, 4 (2014), 1432–1462.

Educational Data Mining (EDM) Process



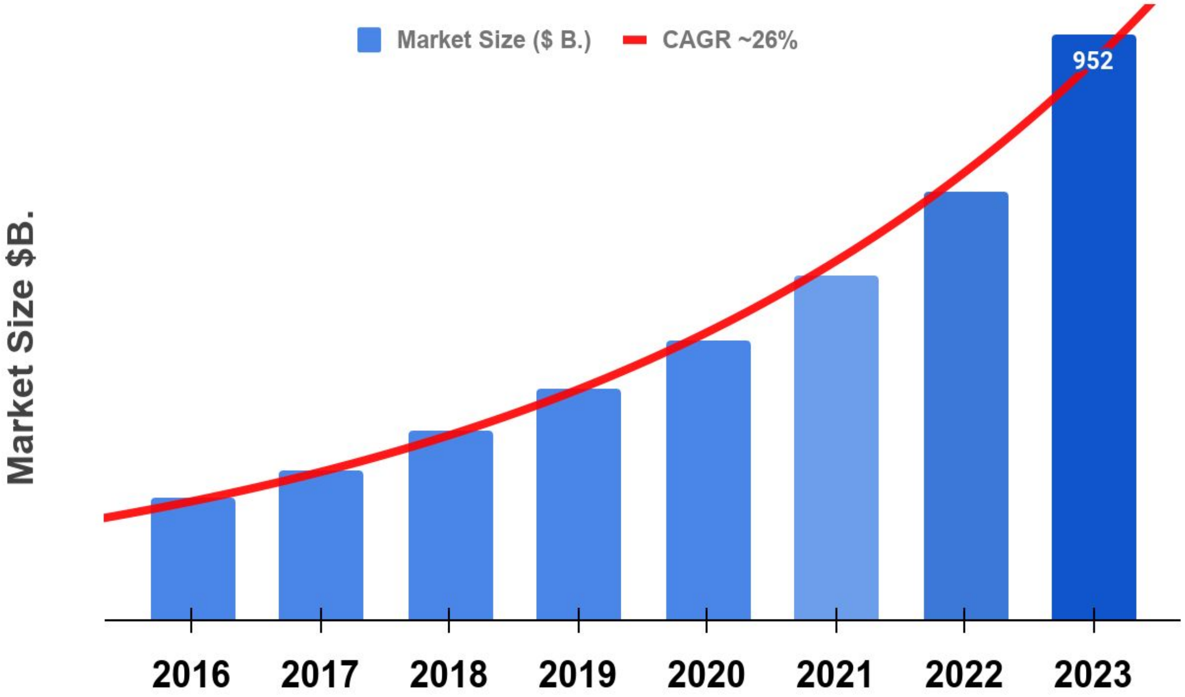
Inspired by https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1355

Educational Data Mining (EDM)

Help Answer Question Like:

- What sequence of **topics** is **most effective** for a specific student?
- Which student **action** are **associated** with better learning and **higher grades**?
- Which **Actions** indicate satisfaction and **engagement**?
- What **features** of an online **learning environment** lead to better learning?
- Which **student** is more **prone** to **drop** studies?

Learning Market



Compound annual growth rate (**CAGR**) is the **rate of return** that would be **required** for an investment to **grow** from its **beginning** balance to its ending **balance**.

Smart Education and Learning Market

Software:

- Learning Management System (LMS)
- Learning Content Management System
- Adaptive Learning Platform
- Assessment Services
- Others

Hardware:

- Interactive White Boards (WBS)
- Interactive Displays
- Interactive Tables
- Student Response Systems

Service:

- Managed Services
- Professional Services

Application:

- Government
- Enterprise/Business Education
- NGOs and Association
- Professional Services
- Healthcare

Deployment:

- Cloud
- On-Premise

Organization Size:

- Small And Medium Organization
- Large Organization

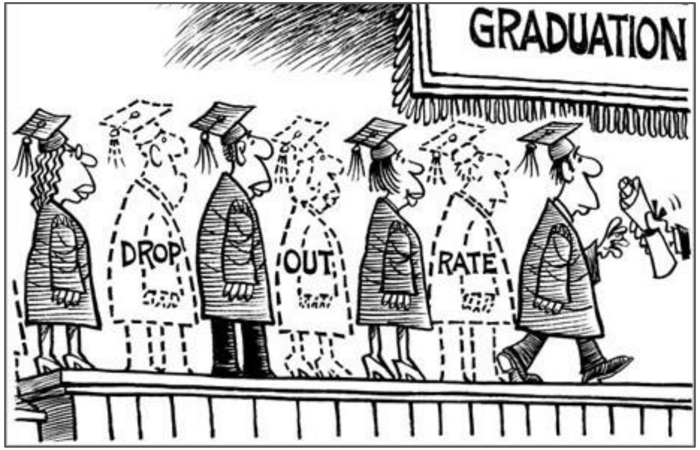
Dropout in Online Environment

- Students in **online degree** programs have a **higher chance** of dropping out than those attending a **conventional** classroom **environment** [12, 21, 23, 26, 35].

- Smith [68] highlights that **40-80%** of **online** students **drop out** from online **classes**

- Their **retention rate** is approximately **10-20% lower** than that of **traditional** universities [34].

- Possibliy because, **students** can **leave** the course at any time **without notice** and further consequences.



[12] Vicki Carter. 1996. Do media influence learning? Revisiting the debate in the context of distance education. Open Learning: The Journal of Open, Distance and e-Learning 11, 1 (1996), 31–40.
[21] David P. Diaz. 2000. Comparison of student characteristics, and evaluation of student success, in an online health education course. Ph.D. Dissertation. Nova Southeastern University.
[23] William Doherty. 2006. An analysis of multiple factors affecting retention in web-based community college courses. The Internet and Higher Education 9, 4 (2006), 245–255.
[26] Karen Frankola.2 001. Why online learners dropout. WORKFORCE-COSTA MESA-80,10(2001),52–61. http://www.workforce.com/feature/00/07/29
[34] Michael Herbert. 2006. Staying the course: A study in online student satisfaction and retention. Online Journal of Distance Learning Administration 9, 4 (2006), 300–317.
[35] Erin Heyman. 2010. Overcoming student retention issues in higher education online programs. Online Journal of Distance Learning Administration 13, 4 (2010).
[68] Belinda G. Smith. 2010. E-learning technologies: A comparative study of adult learners enrolled on blended and online campuses engaging in a virtual classroom. Ph.D. Dissertation. Capella University.
[70] Denise E. Stanford-Bowers. 2008. Persistence in online classes: A study of perceptions among community college stakeholders. Journal of Online Learning and Teaching 4, 1 (2008), 37–50.

Benefits of SDP strategies

- **direct** impact: by **increasing** the **retention** and **completion** rates;

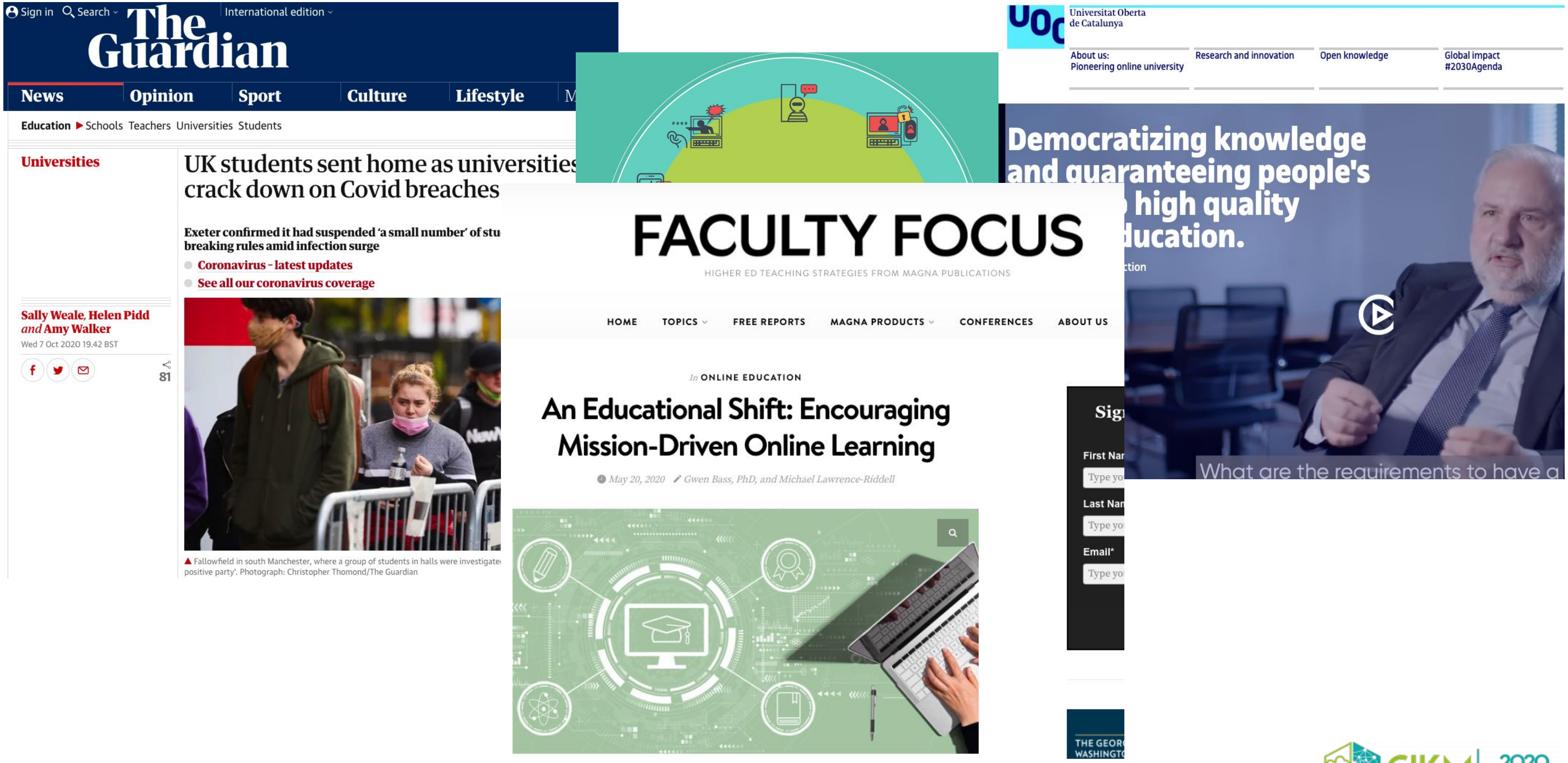
- improve **learning quality**: helps to **develop** intervention **strategies** to provide individually **tailored** support;

- dropouts cause significant **economic wastes**: universities have a clear interest in **investing** in this type of **predictive actions**;

- **prestige point-of-view:** institutions who exhibit **higher** graduation **rates attract** a higher number of **students**;

[12] Vicki Carter. 1996. Do media influence learning? Revisiting the debate in the context of distance education. Open Learning: The Journal of Open, Distance and e-Learning 11, 1 (1996), 31–40.
[21] David P. Diaz. 2000. Comparison of student characteristics, and evaluation of student success, in an online health education course. Ph.D. Dissertation. Nova Southeastern University.
[23] William Doherty. 2006. An analysis of multiple factors affecting retention in web-based community college courses. The Internet and Higher Education 9, 4 (2006), 245–255.
[26] Karen Frankola.2 001.Why online learners dropout. WORKFORCE-COSTA MESA-80,10(2001),52–61. http://www.workforce.com/feature/00/07/29
[34] Michael Herbert. 2006. Staying the course: A study in online student satisfaction and retention. Online Journal of Distance Learning Administration 9, 4 (2006), 300–317.
[35] Erin Heyman. 2010. Overcoming student retention issues in higher education online programs. Online Journal of Distance Learning Administration 13, 4 (2010).
[70] Denise E. Stanford-Bowers. 2008. Persistence in online classes: A study of perceptions among community college stakeholders. Journal of Online Learning and Teaching 4, 1 (2008), 37–50.

Shift to an online environment



Sign in Search The Guardian International edition

News Opinion Sport Culture Lifestyle M

Education ▶ Schools Teachers Universities Students

Universities

UK students sent home as universities crack down on Covid breaches

Exeter confirmed it had suspended 'a small number' of stu breaking rules amid infection surge

- Coronavirus - latest updates
- See all our coronavirus coverage

Sally Weale, Helen Pidd *and* Amy Walker

Wed 7 Oct 2020 19.42 BST

81

▲ Fallowfield in south Manchester, where a group of students in halls were investigate positive party'. Photograph: Christopher Thomond/The Guardian

UOC Universitat Oberta de Catalunya

About us: Pioneering online university Research and innovation Open knowledge Global impact #2030Agenda

Democratizing knowledge and guaranteeing people's high quality ducation.

ction

What are the requirements to have a

FACULTY FOCUS

HIGHER ED TEACHING STRATEGIES FROM MAGNA PUBLICATIONS

HOME TOPICS FREE REPORTS MAGNA PRODUCTS CONFERENCES ABOUT US

In ONLINE EDUCATION

An Educational Shift: Encouraging Mission-Driven Online Learning

May 20, 2020 *Gwen Bass, PhD, and Michael Lawrence-Riddell*

Sig

First Nar

Type yo

Last Nar

Type yo

Email*

Type yo

THE GEORG WASHINGT

The five dimensions to study SDP

- Prenkaj et al. [61] describes the five dimensions of the **principal literature reviews** today available in the **field** of **SDP**:
    - **Field of Study**, which affects the perspective and objectives of the study;

      Analytic / Computational
    - **Gathered Data** used to analyse the problem;

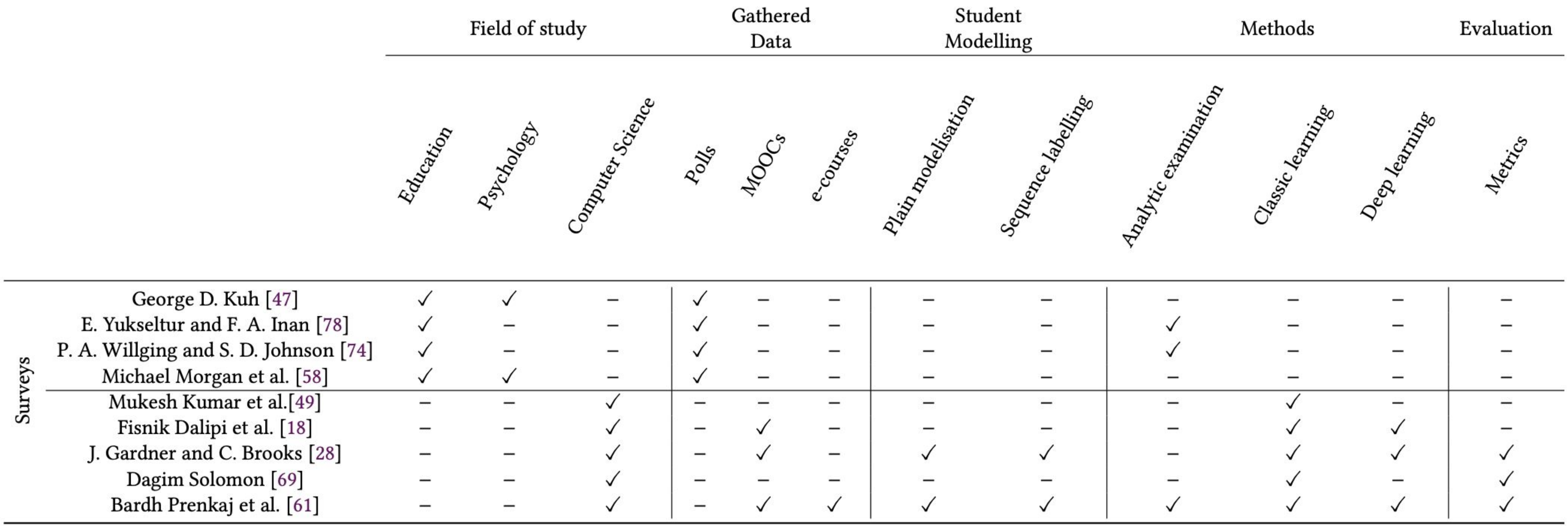
      Polls and questionnaires / MOOCs and e-Courses
    - **Student Modelling** strategies employed to process the raw data;

      plain modelisation / sequence labelling
    - **Methods** to model and solve the SDP problem;

      Analytic / machine learning / deep Learning
    - **Evaluation measures.**

[61] Bardh Prenkaj, Paola Velardi, Giovanni Stilo, Damiano Distante, and Stefano Faralli. 2020. A Survey of Machine Learning Approaches for Student Dropout Prediction in Online Courses. 53, 3, Article 57 (May 2020).

Comparison of surveys in the literature



		Field of study			Gathered Data			Student Modelling		Methods			Evaluation
		Education	Psychology	Computer Science	Polls	MOOCs	e-courses	Plain modelisation	Sequence labelling	Analytic examination	Classic learning	Deep learning	Metrics
Surveys	George D. Kuh [47]	✓	✓	–	✓	–	–	–	–	–	–	–	–
	E. Yukseltur and F. A. Inan [78]	✓	–	–	✓	–	–	–	–	✓	–	–	–
	P. A. Willging and S. D. Johnson [74]	✓	–	–	✓	–	–	–	–	✓	–	–	–
	Michael Morgan et al. [58]	✓	✓	–	✓	–	–	–	–	–	–	–	–
	Mukesh Kumar et al.[49]	–	–	✓	–	–	–	–	–	–	✓	–	–
	Fisnik Dalipi et al. [18]	–	–	✓	–	✓	–	–	–	–	✓	✓	–
	J. Gardner and C. Brooks [28]	–	–	✓	–	✓	–	✓	✓	–	✓	✓	✓
	Dagim Solomon [69]	–	–	✓	–	–	–	–	–	–	✓	–	✓
	Bardh Prenkaj et al. [61]	–	–	✓	–	✓	✓	✓	✓	✓	✓	✓	✓

[47] George D. Kuh. 2009. The National Survey of Student Engagement: Conceptual and empirical foundations. New Directions for Institutional Research 2009 (12 2009), 5–20. https://doi.org/10.1002/ir.283

[78] Eran Yukseltur and Fethi Ahmet Inan. 2006. Examining the Factors Affecting Student Dropout in an Online Learning Environment. Turkish Online Journal of Distance Education 7, 3 (2006), 76–88.

[74] Pedro A. Willging and Scott D. Johnson. 2009. Factors that influence students' decision to dropout of online courses. Journal of Asynchronous Learning Networks 13, 3 (2009), 115–127.

[58] Michael Morgan, Matthew Butler, Neena Thota, and Jane Sinclair. 2018. How CS academics view student engagement. In Proceedings of the 23rd Annual ACM Conference on Innovation and Tech in Computer Science Education. ACM

[49] Mukesh Kumar, AJ Singh, and Disha Handa. 2017. Literature survey on educational dropout prediction. International Journal of Education and Management Engineering 7, 2 (2017), 8.

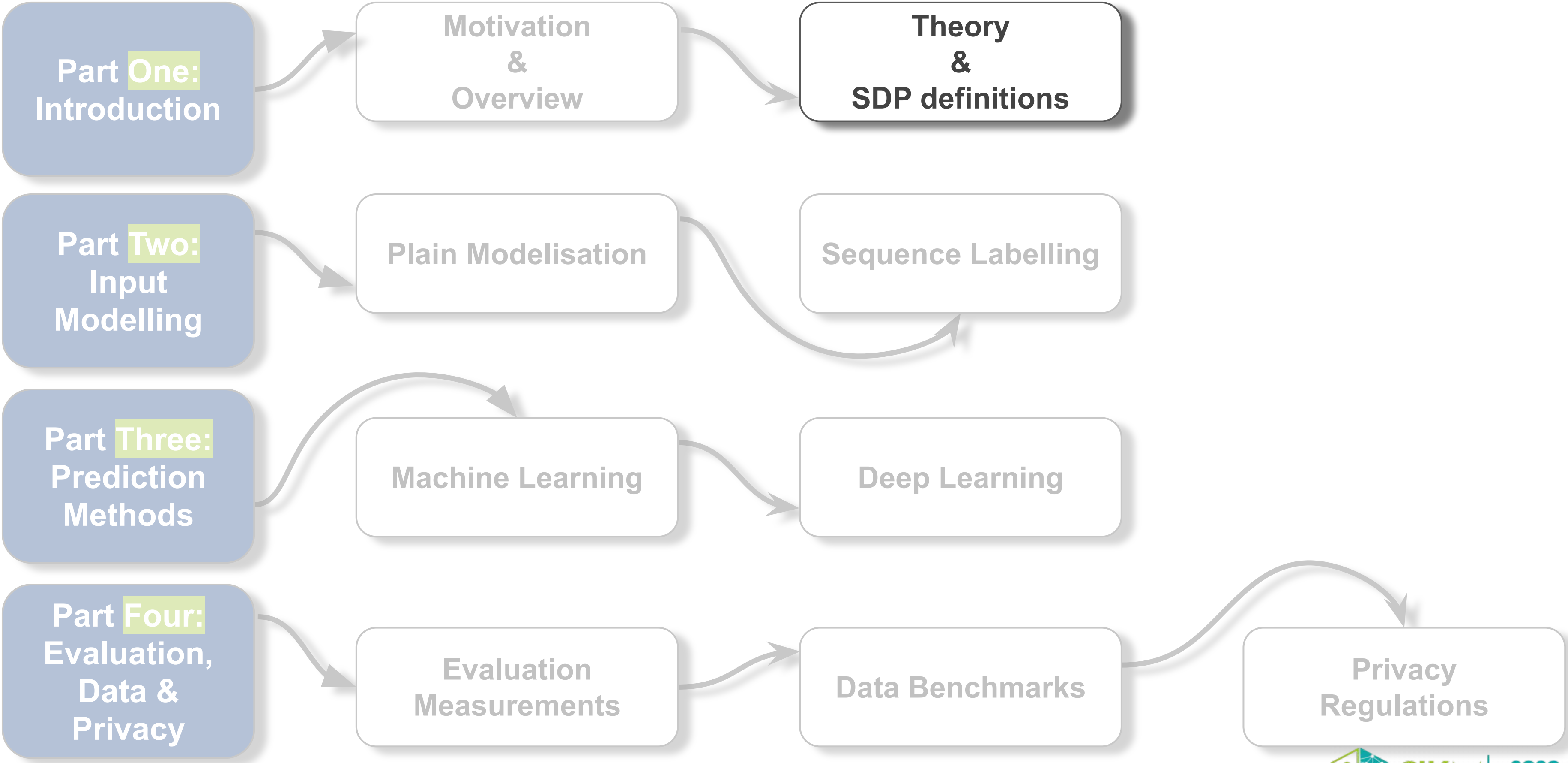
[18] Fisnik Dalipi, Ali Shariq Imran, and Zenun Kastrati. 2018. MOOC dropout prediction using machine learning techniques: Review and research challenges. In 2018 IEEE Global Engineering Education Conference (EDUCON). IEEE, 1007–101

[28] Josh Gardner and Christopher Brooks. 2018. Student success prediction in MOOCs.User Modeling and User-Adapted Interaction 28,2(2018),127–203.

[69] Dagim Solomon. 2018. Predicting Performance and Potential Difficulties of University Student using Classification: Survey Paper. International Journal of Pure and Applied Mathematics 118, 18 (2018), 2703–2707.

[61] Bardh Prenkaj, Paola Velardi, Giovanni Stilo, Damiano Distante, and Stefano Faralli. 2020. A Survey of Machine Learning Approaches for Student Dropout Prediction in Online Courses. 53, 3, Article 57 (May 2020).

Roadmap



Part One: Introduction

Motivation & Overview

Theory & SDP definitions

Part Two: Input Modelling

Plain Modelisation

Sequence Labelling

Part Three: Prediction Methods

Machine Learning

Deep Learning

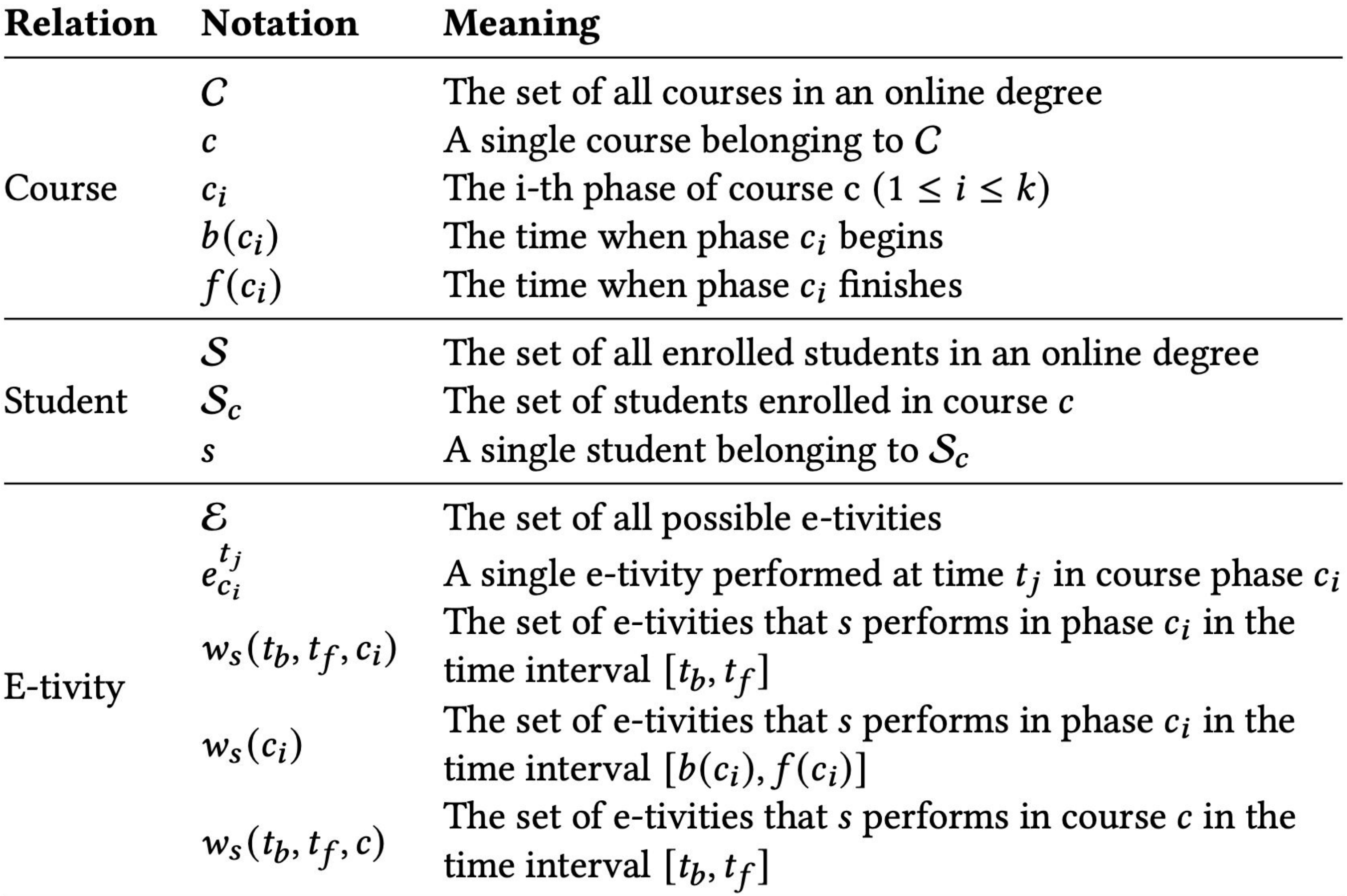
Part Four: Evaluation, Data & Privacy

Evaluation Measurements

Data Benchmarks

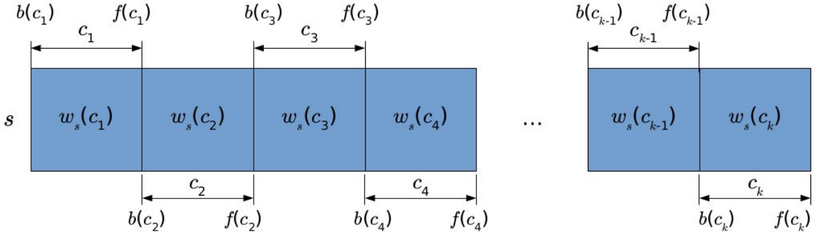
Privacy Regulations

Adopted Notation



Relation	Notation	Meaning
Course	$\mathcal{C}$	The set of all courses in an online degree
	c	A single course belonging to $\mathcal{C}$
	c_i	The i-th phase of course c $(1 \leq i \leq k)$
	$b(c_i)$	The time when phase c_i begins
	$f(c_i)$	The time when phase c_i finishes
Student	$\mathcal{S}$	The set of all enrolled students in an online degree
	$\mathcal{S}_c$	The set of students enrolled in course c
	s	A single student belonging to $\mathcal{S}_c$
E-tivity	$\mathcal{E}$	The set of all possible e-tivities
	$e_{c_i}^{t_j}$	A single e-tivity performed at time t_j in course phase c_i
	$w_s(t_b, t_f, c_i)$	The set of e-tivities that s performs in phase c_i in the time interval $[t_b, t_f]$
	$w_s(c_i)$	The set of e-tivities that s performs in phase c_i in the time interval $[b(c_i), f(c_i)]$
	$w_s(t_b, t_f, c)$	The set of e-tivities that s performs in course c in the time interval $[t_b, t_f]$

SDP’s Formulations



- ***Plain dropout formulation:*** The student-platform **interactions** are **independent** in time. Given the e-tivities $w_s(t_b,t_f,c)$ that $s \in S_c$ performs, **plain dropout** determines whether s **drops out or not** regardless of how the e-tivities are sequenced in the interval $[t_b,t_f]$.

- ***Recurrent dropout formulation:*** Uses information from **previous** course **phases** to decide a student’s dropout status. Hidden information from phase c_{i-1} to c_{i}.
  Therefore, the **label** of s in phase c_i **depends** on the **activities** performed in the **preceding** phases c_{i-r} , $r \in \{1, ..., p \leq i - 1\}$.
  p is the the window used to consider previous phases.

SDP’s Definitions[1]

- **Definition 1. Plain dropout:** *Given a time interval [t_b,t_f], a student s is a dropout from course c if they do not survive until the end of the time span. In other words, s is a dropout if ∃ t_u ∈ [t_b,t_f] s.t. w_s (t_u,t_f ,c) = ∅.*

- The dropout condition does not depend on the information passed from phase c_{i-1} to c_i .

- The dropout label is based on all the e-tivities w_s (t_u,t_f,c) that s does after a certain point in time t_u.

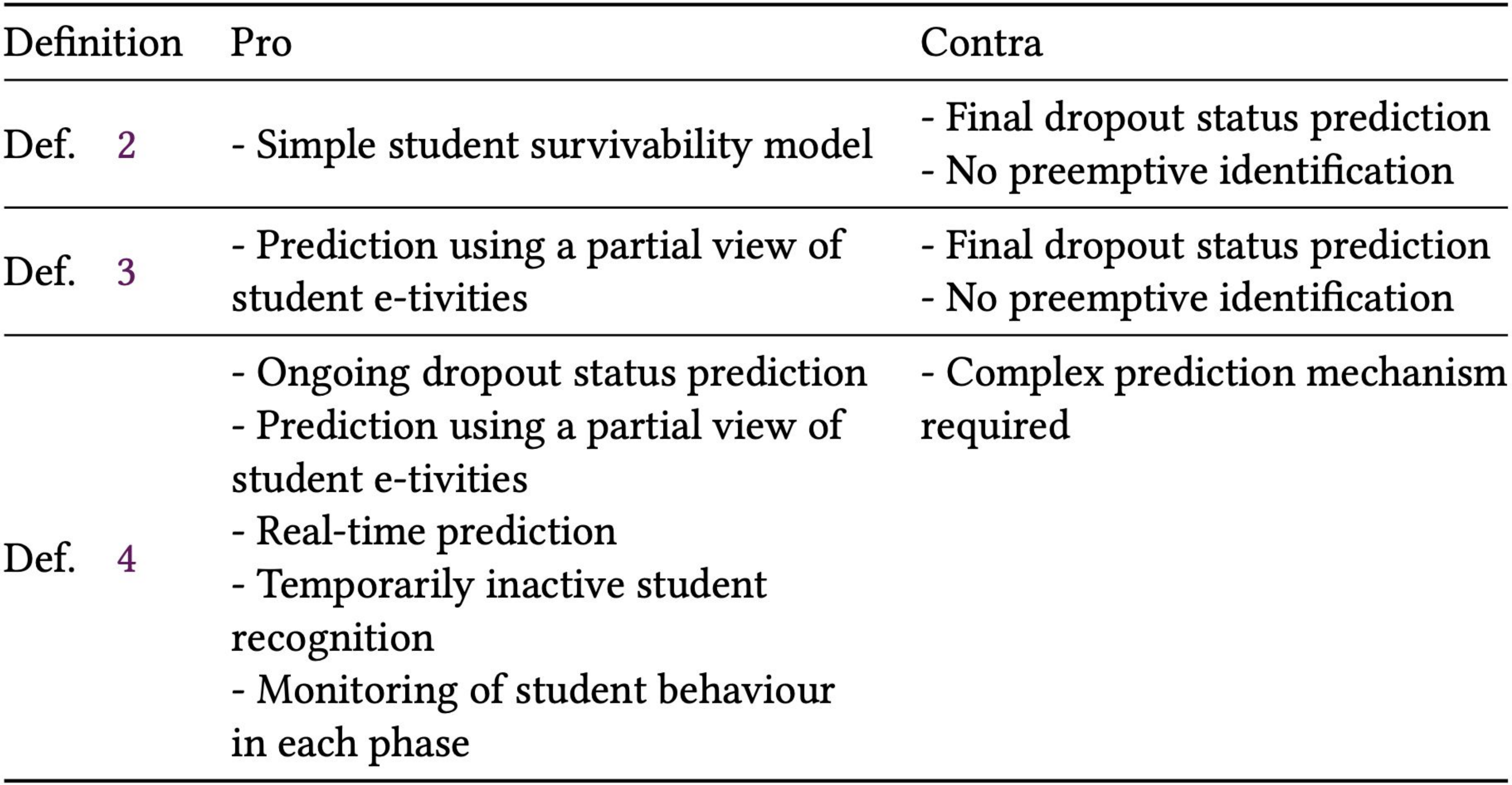
SDP's Definitions [2]

- **Definition 2. Participation in the final course phase:** *A student s is a dropout if they do not **persist** until the **last phase**, c_k , of a course c; otherwise, they are a persister. In other words, s is a dropout if w_s (c_k) = ∅.*

- **Definition 3. Last phase of engagement:**
  A student s is a dropout if they do not produce any e-tivities after the current phase c_i: i.e. s is a dropout if $w_s(c_i) \neq \varnothing \land \forall j \in [i+1,k]$, $w_s(c_j) = \varnothing$.
  Notice that this definition is a generalisation of the previous one:
  i.e. we emulate 2 by setting $i = k - 1$.

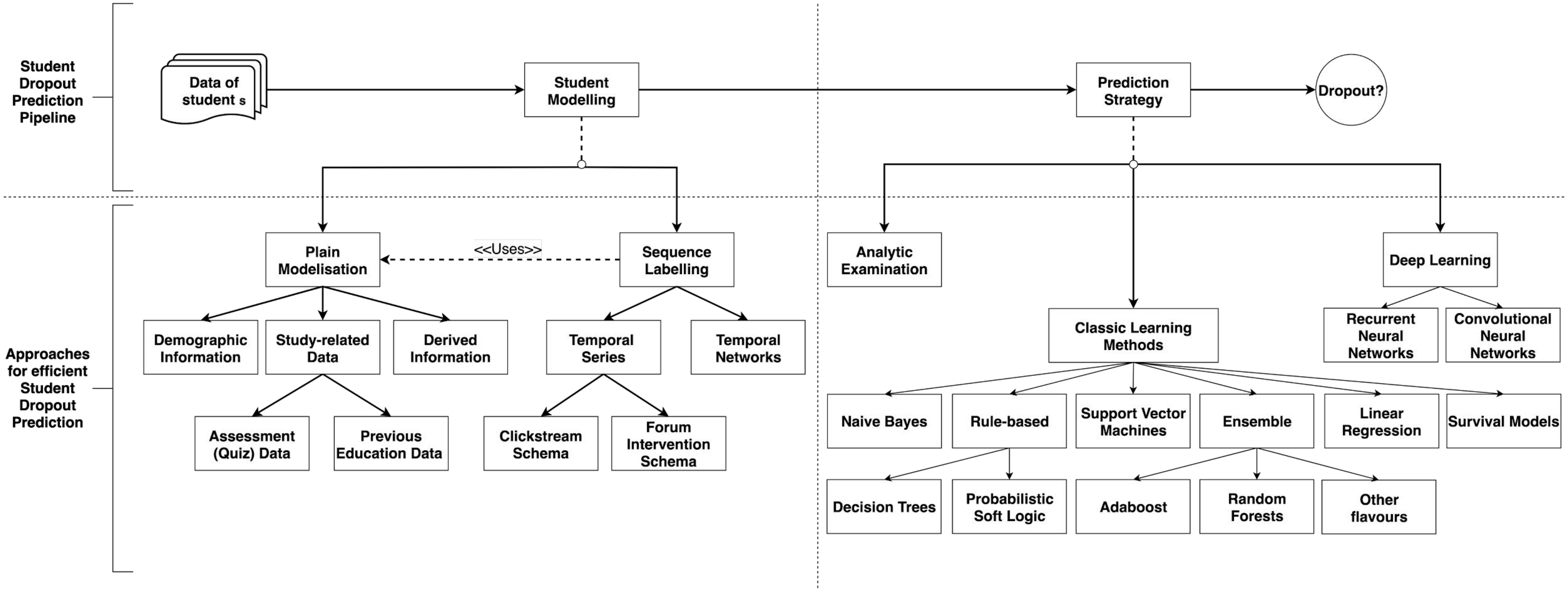
- **Definition 4. Participation in the next phase:**
  A student s is a dropout if they do not have e-tivities in the next phase, c_i+1.
  Hence,s is a dropout if $w_s(c_i+1) = \varnothing \land i \neq k$.

Advantages and disadvantages

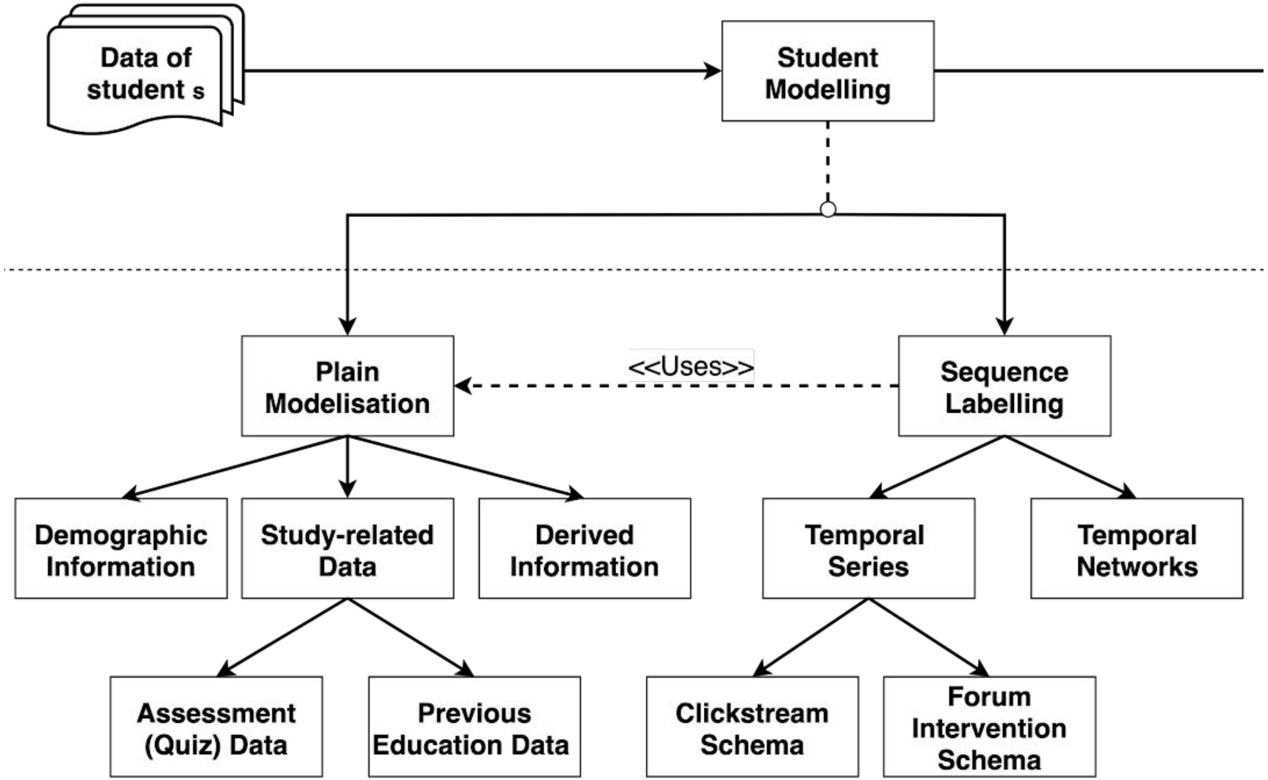


Definition	Pro	Contra
Def. 2	- Simple student survivability model	- Final dropout status prediction - No preemptive identification
Def. 3	- Prediction using a partial view of student e-tivities	- Final dropout status prediction - No preemptive identification
Def. 4	- Ongoing dropout status prediction - Prediction using a partial view of student e-tivities - Real-time prediction - Temporarily inactive student recognition - Monitoring of student behaviour in each phase	- Complex prediction mechanism required

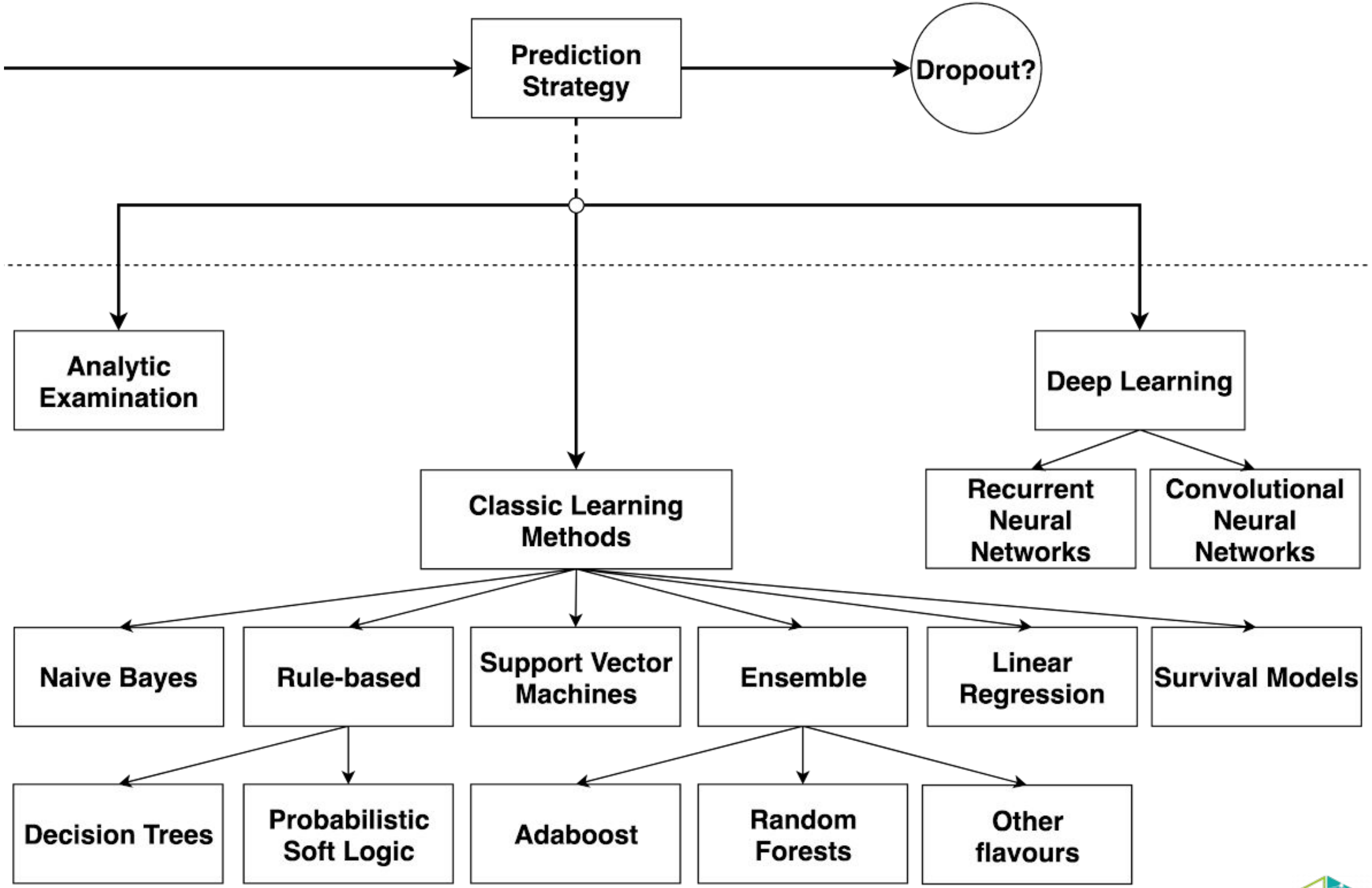
A taxonomy of SDP design choices



A taxonomy of student modelling approaches



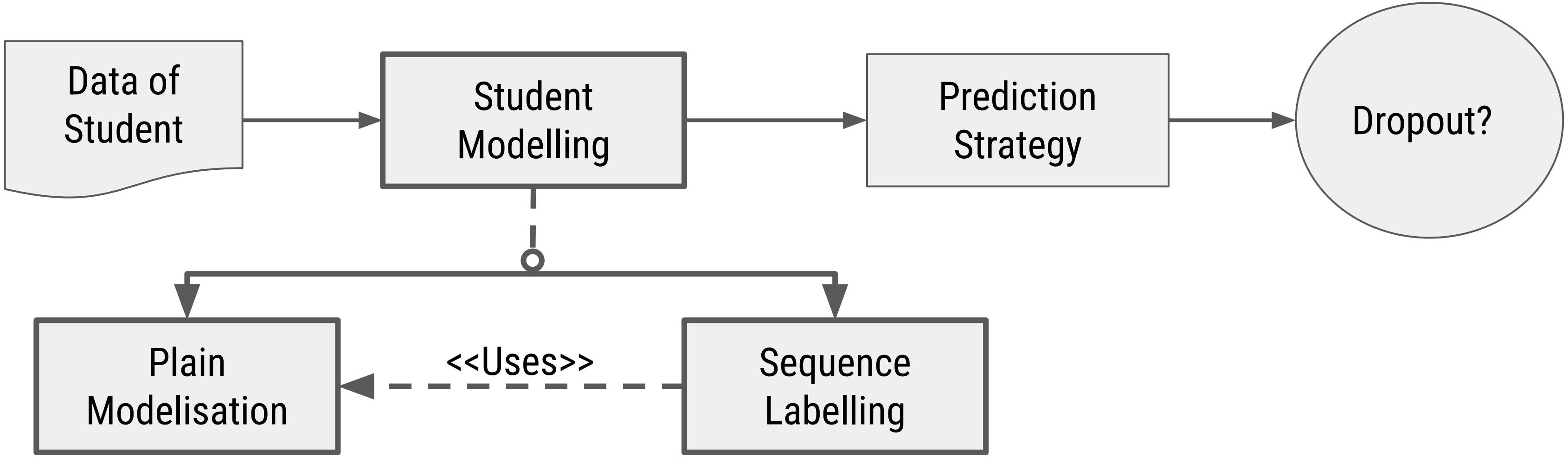
A taxonomy of prediction strategy approaches



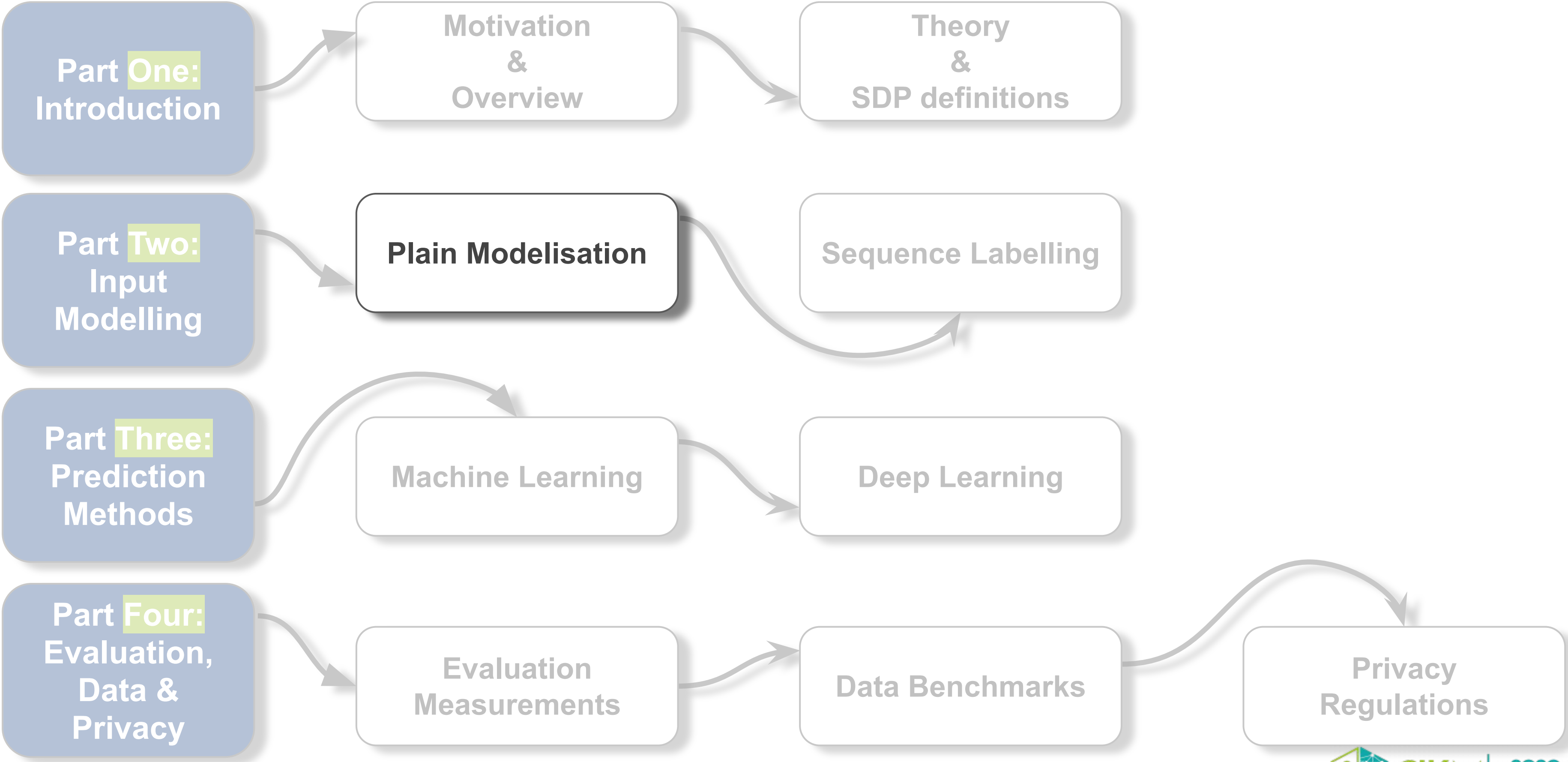
Part Two
Input modelling techniques

Student Data Modelling Techniques

- Plain modelisation
- Sequence labelling



Roadmap



Part One: Introduction

Motivation & Overview

Theory & SDP definitions

Part Two: Input Modelling

Plain Modelisation

Sequence Labelling

Part Three: Prediction Methods

Machine Learning

Deep Learning

Part Four: Evaluation, Data & Privacy

Evaluation Measurements

Data Benchmarks

Privacy Regulations

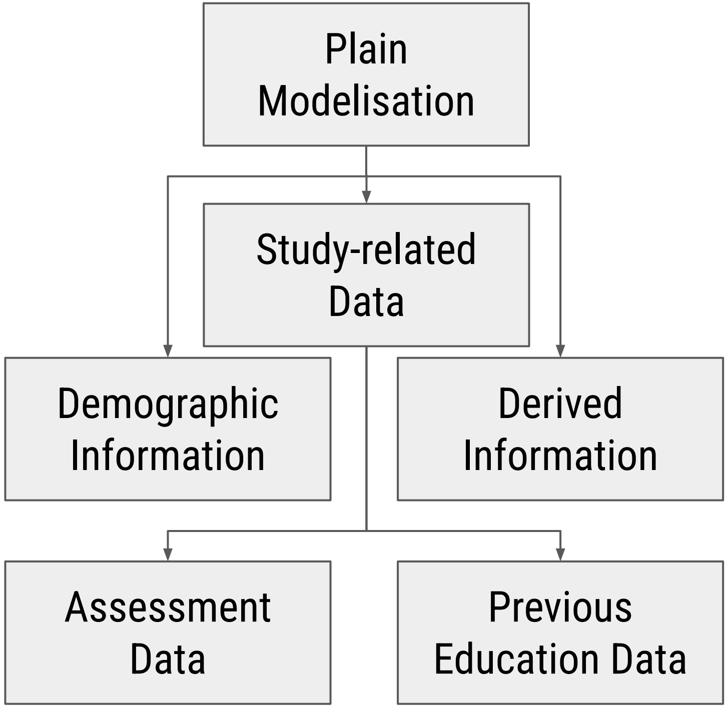
Plain Modelisation

Plain modelisation exploits student demographic information and study-related data as input for a prediction strategy.

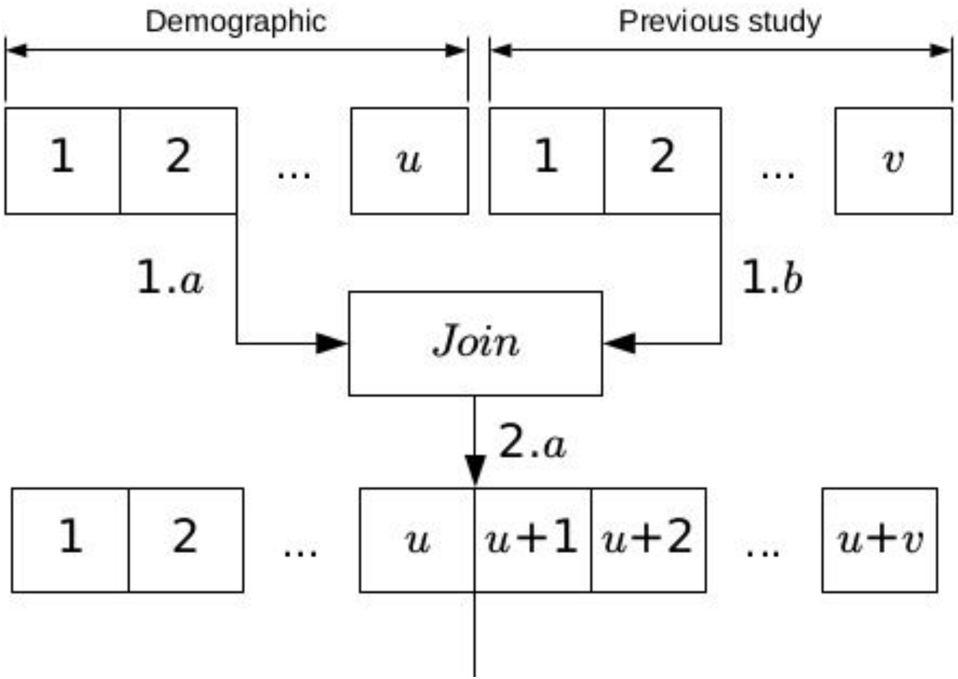
- Static or time-invariant data
- Flatten time-variant data

Plain Modelisation

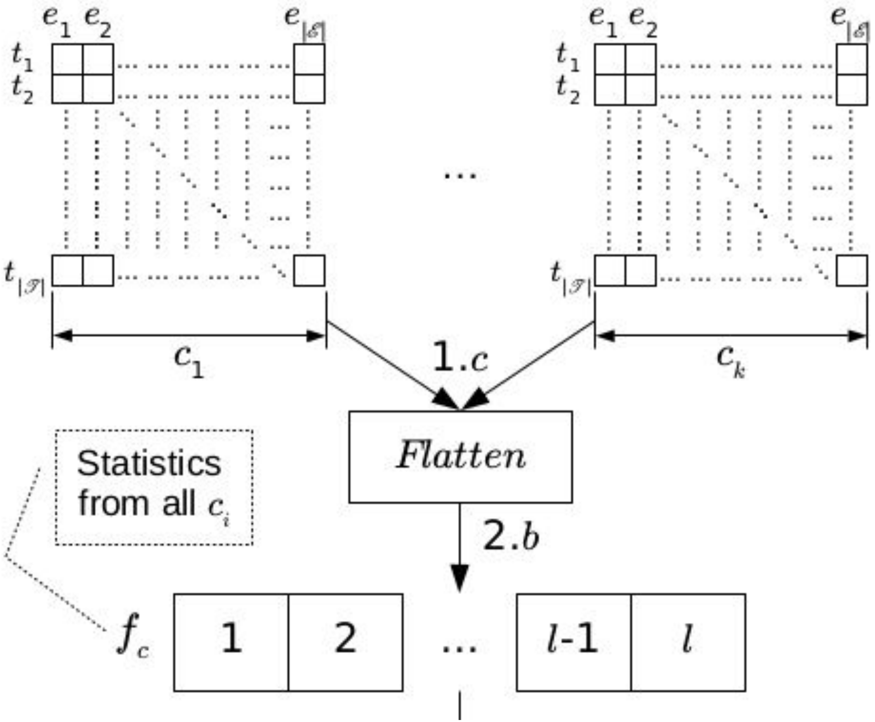
- Demographic Information
- Study-related Data
  - Assessment (Quiz) Data
  - Previous Education Data
- Derived Information



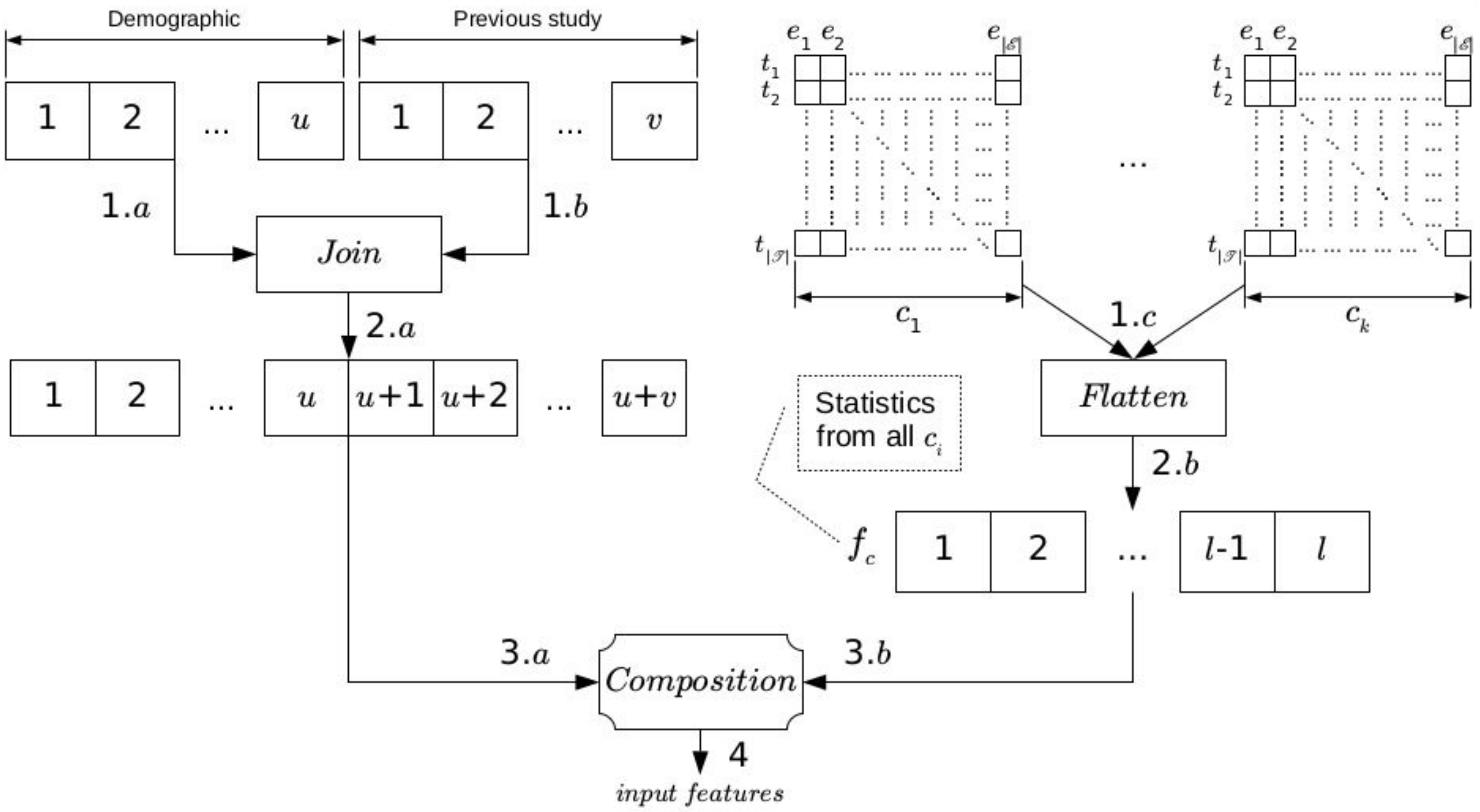
Plain Modelisation



Plain Modelisation



Plain Modelisation



Demographic Information

Demographic information describes the living context of the student:

- Personal characteristics
- Economic-related information
- Family-related information

Demographic Information

- Personal characteristics:
  - Sex[1, 2, 8, 42, 44, 45, 46, 53, 76]
  - Age[1, 8, 42, 44, 45, 46,75]
  - Ethnicity[2, 46]
  - Residency[1, 2, 53]
  - Citizenship[8]
  - English language skill level[53]
  - Computer literacy[42, 44, 45, 76]
  - Disability[46]

[1] Qasem A. Al-Radaideh, Emad M. Al-Shawakfa, and Mustafa I. Al-Najjar. 2006. Mining student data using decision trees. In Proceedings of the International Arab Conference on Information Technology (ACIT'2006). 1–5.

[8] Johannes Berens, Kerstin Schneider, Simon Görtz, Simon Oster, and Julian Burghoff. 2019. Early detection of students at risk - Predicting student dropouts using administrative student data from German Universities and machine learning methods. Journal of Educational Data Mining 11, 3 (2019), 1–41.

[42] Georgios Kostopoulos, Sotiris Kotsiantis, and Panagiotis Pintelas. 2015. Estimating student dropout in distance higher education using semi-supervised techniques. In Proceedings of the 19th Panhellenic Conference on Informatics. ACM, New York, NY, 38–43.

[44] Sotiris Kotsiantis, Christos Pierrakeas, and Panagiotis Pintelas. 2003. Preventing student dropout in distance learning using machine learning techniques. In Proceedings of the International Conference on Knowledge-based and Intelligent Information and Engineering Systems. Springer, New York, NY, 267–274.

[45] Sotiris Kotsiantis, Christos Pierrakeas, Ioannis Zaharakis, and Panagiotis Pintelas. 2003. Efficiency of Machine Learning Techniques in Predicting Students Performance in Distance Learning Systems. University of Patras Press, 297–306.

[46] Zlatko J. Kovačić. 2010. Early prediction of student success: Mining student enrollment data. In Proceedings of the Informing Science & IT Education Conference. Citeseer.

[53] Ioanna Lykourentzou, Ioannis Giannoukos, Vassilis Nikolopoulos, George Mpardis, and Vassili Loumos. 2009. Dropout prediction in e-learning courses through the combination of machine learning techniques. Comput. Educ. 53, 3 (2009), 950–965.

[76] Michalis Xenos, Christos Pierrakeas, and Panagiotis Pintelas. 2002. A survey on student dropout rates and dropout causes concerning the students in the course of informatics of the Hellenic Open University. Comput. Educ. 39, 4 (2002), 361–377.

Demographic Information

- Economic-related information:
    - ○ Occupation[42, 44, 45, 46, 53]
    - ○ Funding[1]

[1] Qasem A. Al-Radaideh, Emad M. Al-Shawakfa, and Mustafa I. Al-Najjar. 2006. Mining student data using decision trees. In Proceedings of the International Arab Conference on Information Technology (ACIT'2006). 1–5.
[42] Georgios Kostopoulos, Sotiris Kotsiantis, and Panagiotis Pintelas. 2015. Estimating student dropout in distance higher education using semi-supervised techniques. In Proceedings of the 19th Panhellenic Conference on Informatics. ACM, New York, NY, 38–43.
[44] Sotiris Kotsiantis, Christos Pierrakeas, and Panagiotis Pintelas. 2003. Preventing student dropout in distance learning using machine learning techniques. In Proceedings of the International Conference on Knowledge-based and Intelligent Information and Engineering Systems. Springer, New York, NY, 267–274.
[45] Sotiris Kotsiantis, Christos Pierrakeas, Ioannis Zaharakis, and Panagiotis Pintelas. 2003. Efficiency of Machine Learning Techniques in Predicting Students Performance in Distance Learning Systems. University of Patras Press, 297–306.
[46] Zlatko J. Kovačić. 2010. Early prediction of student success: Mining student enrollment data. In Proceedings of the Informing Science & IT Education Conference. Citeseer.
[53] Ioanna Lykourentzou, Ioannis Giannoukos, Vassilis Nikolopoulos, George Mpardis, and Vassili Loumos. 2009. Dropout prediction in e-learning courses through the combination of machine learning techniques. Comput. Educ. 53, 3 (2009), 950–965.

Demographic Information

- Family information:
  - Marital Status[2, 42, 44, 45, 76]
  - Number of children[42, 44, 45]
  - Parents education[2]
  - Parents occupation[2]

[2] Sattar Ameri, Mahtab J. Fard, Ratna B. Chinnam, and Chandan K. Reddy. 2016. Survival analysis based framework for early prediction of student dropouts. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 903–912.
[42] Georgios Kostopoulos, Sotiris Kotsiantis, and Panagiotis Pintelas. 2015. Estimating student dropout in distance higher education using semi-supervised techniques. In Proceedings of the 19th Panhellenic Conference on Informatics. ACM, New York, NY, 38–43.
[44] Sotiris Kotsiantis, Christos Pierrakeas, and Panagiotis Pintelas. 2003. Preventing student dropout in distance learning using machine learning techniques. In Proceedings of the International Conference on Knowledge-based and Intelligent Information and Engineering Systems. Springer, New York, NY, 267–274.
[45] Sotiris Kotsiantis, Christos Pierrakeas, Ioannis Zaharakis, and Panagiotis Pintelas. 2003. Efficiency of Machine Learning Techniques in Predicting Students Performance in Distance Learning Systems. University of Patras Press, 297–306.
[76] Michalis Xenos, Christos Pierrakeas, and Panagiotis Pintelas. 2002. A survey on student dropout rates and dropout causes concerning the students in the course of informatics of the Hellenic Open University. Comput. Educ. 39, 4 (2002), 361–377.

Study-related Data

Study-related data provides a view of the effort and interests of the student.

- Current study-related data
- Assessment (Quiz) Data
- Previous Education Data

Study-related Data

- Current study-related features:
  - Enrollment of the student[46]
  - Form of study[1]
  - Lecturer’s working position, sex[1]

[1] Qasem A. Al-Radaideh, Emad M. Al-Shawakfa, and Mustafa I. Al-Najjar. 2006. Mining student data using decision trees. In Proceedings of the International Arab Conference on Information Technology (ACIT’2006). 1–5.
[46] Zlatko J. Kovačić. 2010. Early prediction of student success: Mining student enrollment data. In Proceedings of the Informing Science & IT Education Conference. Citeseer.

Study-related Data: Assessment Data

- Assessment Features:
  - Grades[8, 29, 42, 43, 44, 45, 53, 75,76]
  - Number of submissions[3, 43]
  - Participation[42, 43, 44, 45, 76]

[3] Bussaba Amnueypornsakul, Suma Bhat, and Phakpoom Chinprutthiwong. 2014. Predicting attrition along the way: The UIUC model. In Proceedings of the EMNLP Workshop on Analysis of Large Scale Social Interaction in MOOCs. Association for Computational Linguistics, 55–59.
[8] Johannes Berens, Kerstin Schneider, Simon Görtz, Simon Oster, and Julian Burghoff. 2019. Early detection of students at risk - Predicting student dropouts using administrative student data from German Universities and machine learning methods. Journal of Educational Data Mining 11, 3 (2019), 1–41.
[42] Georgios Kostopoulos, Sotiris Kotsiantis, and Panagiotis Pintelas. 2015. Estimating student dropout in distance higher education using semi-supervised techniques. In Proceedings of the 19th Panhellenic Conference on Informatics. ACM, New York, NY, 38–43.
[43] Sotiris Kotsiantis, Kiriakos Patriarcheas, and Michalis Xenos. 2010. A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. Knowl.-based Syst. 23, 6 (2010), 529–535.
[45] Sotiris Kotsiantis, Christos Pierrakeas, Ioannis Zaharakis, and Panagiotis Pintelas. 2003. Efficiency of Machine Learning Techniques in Predicting Students Performance in Distance Learning Systems. University of Patras Press, 297–306.
[53] Ioanna Lykourentzou, Ioannis Giannoukos, Vassilis Nikolopoulos, George Mpardis, and Vassili Loumos. 2009. Dropout prediction in e learning courses through the combination of machine learning techniques. Comput. Educ. 53, 3 (2009), 950–965.
[75] Annika Wolff, Zdenek Zdrahal, Andriy Nikolov, and Michal Pantucek. 2013. Improving retention: Predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In Proceedings of the 3rd International Conference on Learning Analytics and Knowledge. ACM, New York, NY, 145–149.
[76] Michalis Xenos, Christos Pierrakeas, and Panagiotis Pintelas. 2002. A survey on student dropout rates and dropout causes concerning the students in the course of informatics of the Hellenic Open University. Comput. Educ. 39, 4 (2002), 361–377.

Study-related Data: Previous Education Data

Previous education information describes the skills of the student.

- Previous education type (Scientific, etc..)[1, 20]
- High school country[1, 8, 20]
- Academic qualifications and grades[1, 2, 8, 20, 53]

[1] Qasem A. Al-Radaideh, Emad M. Al-Shawakfa, and Mustafa I. Al-Najjar. 2006. Mining student data using decision trees. In Proceedings of the International Arab Conference on Information Technology (ACIT'2006). 1–5.
[8] Johannes Berens, Kerstin Schneider, Simon Görtz, Simon Oster, and Julian Burghoff. 2019. Early detection of students at risk - Predicting student dropouts using administrative student data from German Universities and machine learning methods. Journal of Educational Data Mining 11, 3 (2019), 1–41.
[20] Gerben W. Dekker, Mykola Pechenizkiy, and Jan M. Vleeshouwer. 2009. Predicting students drop out: A case study. In Proceedings of the International Conference on Educational Data Mining (EDM'09).
[53] Ioanna Lykourentzou, Ioannis Giannoukos, Vassilis Nikolopoulos, George Mpardis, and Vassili Loumos. 2009. Dropout prediction in e-learning courses through the combination of machine learning techniques. Comput. Educ. 53, 3 (2009), 950–965.

Derived Information

Student’s performance and behavioural statistics derived from temporal data or analytical tools.

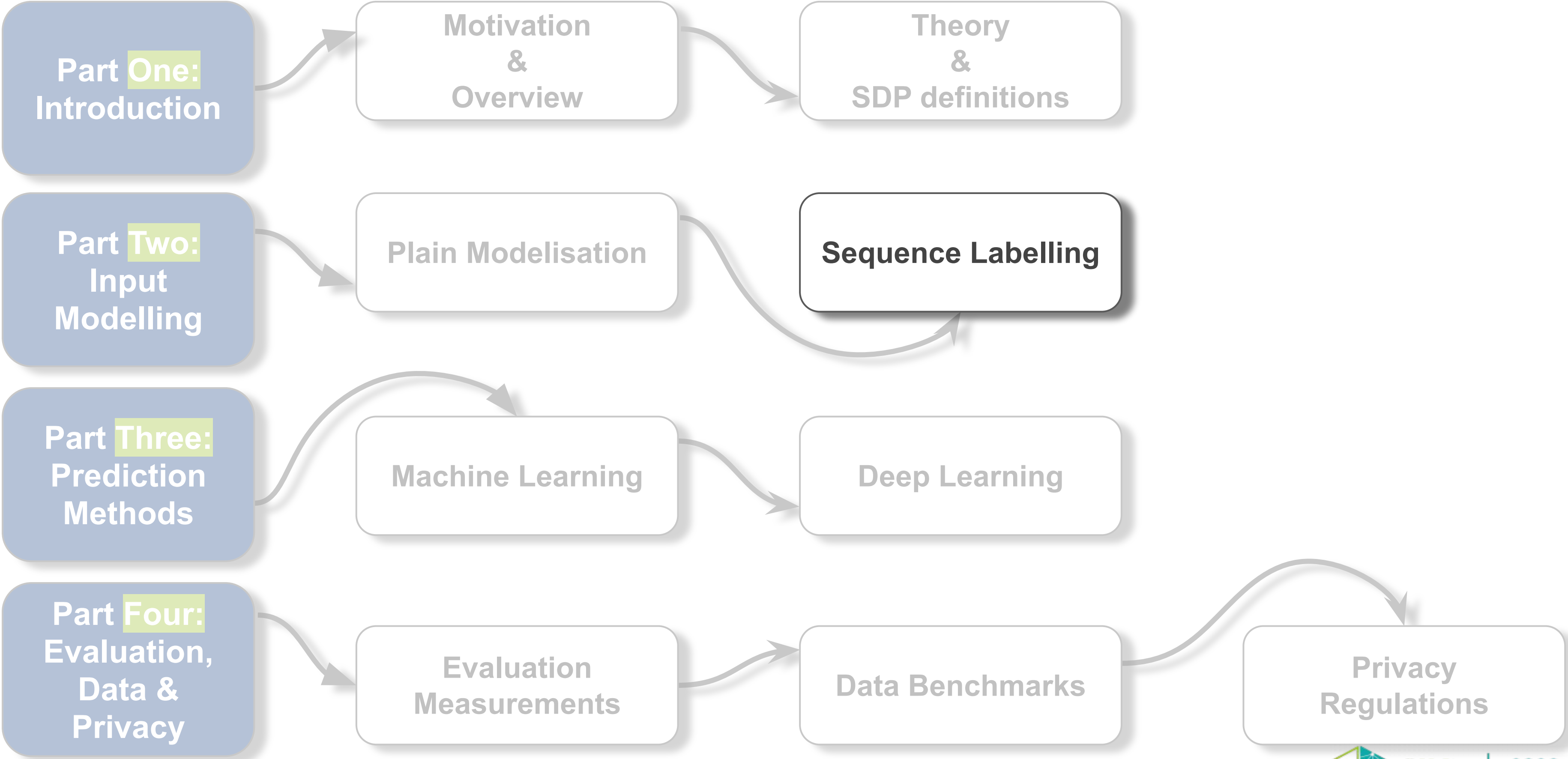
- Average grade, GPA, Earned credits **[2, 8,29, 53, 54]**
- Parallel/Failed/Completed courses[1,8, 31]
- Actions stats(i.e. clicks, page visits, sessions, percentage use a learning resource or forum interactions)[3,25 29, 31, 37, 53, 75]
- NLP stats (word count)[66]

[3] Bussaba Amnueypornsakul, Suma Bhat, and Phakpoom Chinprutthiwong. 2014. Predicting attrition along the way: The UIUC model. In Proceedings of the EMNLP Workshop on Analysis of Large Scale Social Interaction in MOOCs. Association for Computational Linguistics, 55–59.
[25] Wenzheng Feng, Jie Tang, and Tracy Xiao Liu. 2019. Understanding dropouts in MOOCs. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI’19).
[29] Elena Gaudioso, Miguel Montero, and Felix Hernandez-Del-Olmo. 2012. Supporting teachers in adaptive educational systems through predictive models: A proof of concept. Exp. Syst. Applic. 39, 1 (2012), 621–625.
[31] Cameron C. Gray and Dave Perkins. 2019. Utilizing early engagement and machine learning to predict student outcomes. Comput. Educ. 131 (2019), 22–32.
[37] Ya-Han Hu, Chia-Lun Lo, and Sheng-Pao Shih. 2014. Developing early warning systems to predict students’ online learning performance. Comput. Human Behav. 36 (2014), 469–478.
[46] Zlatko J. Kovačić. 2010. Early prediction of student success: Mining student enrollment data. In Proceedings of the Informing Science & IT Education Conference. Citeseer.
[53] Ioanna Lykourentzou, Ioannis Giannoukos, Vassilis Nikolopoulos, George Mpardis, and Vassili Loumos. 2009. Dropout prediction in e-learning courses through the combination of machine learning techniques. Comput. Educ. 53, 3 (2009), 950–965.
[54] Laci Mary Barbosa Manhães, Sérgio Manuel Serra da Cruz, and Geraldo Zimbrão. 2014. WAVE: An architecture for predicting dropout in undergraduate courses using EDM. In Proceedings of the 29th ACM Symposium on Applied Computing. ACM, New York, NY, 243–247.
[66] Carly Robinson, Michael Yeomans, Justin Reich, Chris Hulleman, and Hunter Gehlbach. 2016. Forecasting student achievement in MOOCs with natural language processing. In Proceedings of the 6th International Conference on Learning Analytics & Knowledge. ACM, New York, NY, 383–387.
[75] Annika Wolff, Zdenek Zdrahal, Andriy Nikolov, and Michal Pantucek. 2013. Improving retention: Predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In Proceedings of the 3rd International Conference on Learning Analytics and Knowledge. ACM, New York, NY, 145–149.

Break Time

10 minutes

Roadmap


Part One:
Introduction
Motivation & Overview
Theory & SDP definitions
Part Two:
Input Modelling
Plain Modelisation
Sequence Labelling
Part Three:
Prediction Methods
Machine Learning
Deep Learning
Part Four:
Evaluation, Data & Privacy
Evaluation Measurements
Data Benchmarks
Privacy Regulations

Sequence Labelling

- Discrete-time temporal series
    - Clickstream-based schema
    - Forum-intervention-based schema
- Temporal networks
    - Newcomers respond to all
    - Star networks

Discrete-time temporal series

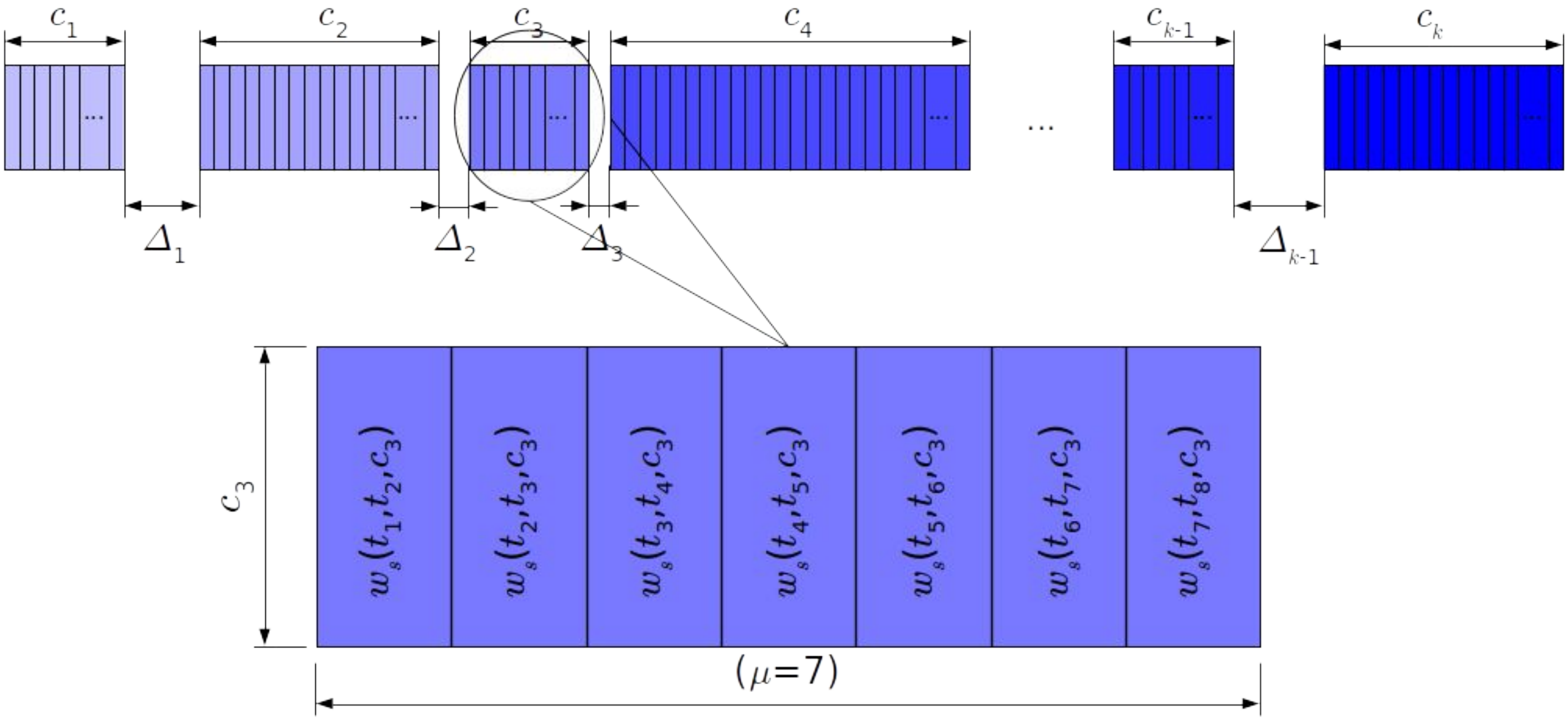
- Shape raw data into *discrete time-series* of e-tivities

- **Discrete time-series** = set of observations recorded at time t belonging to a finite set of times Τ

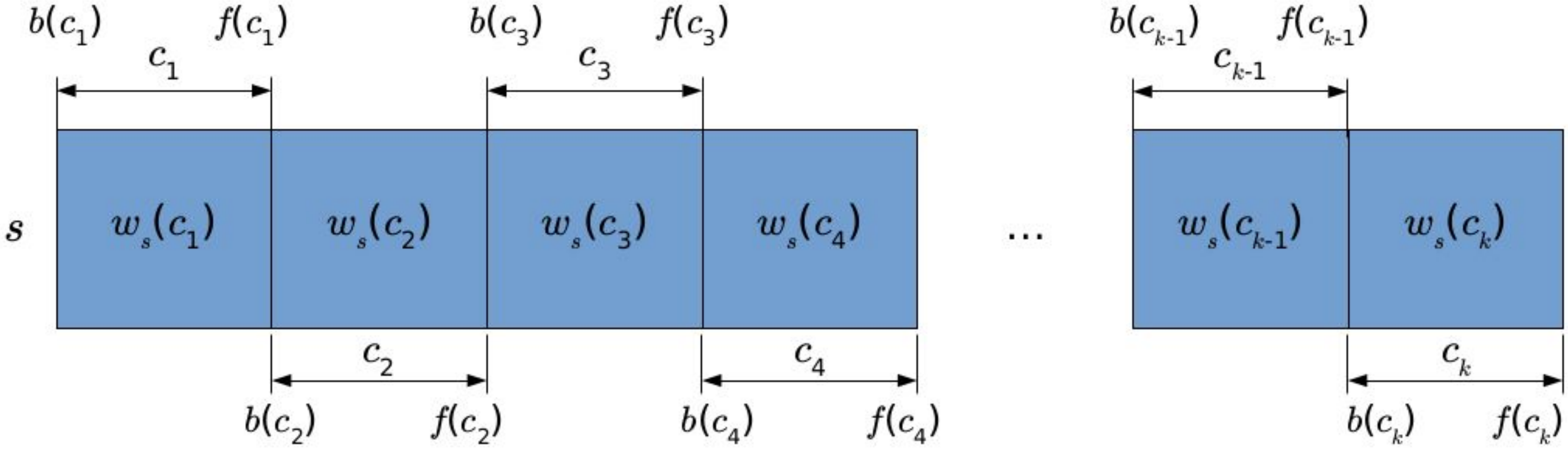
- Divide each course phase into μ consecutive temporal slices

- **Interphasic gaps** or consecutive course phases?

Discrete-time temporal series



Discrete-time temporal series



Clickstream-based schema

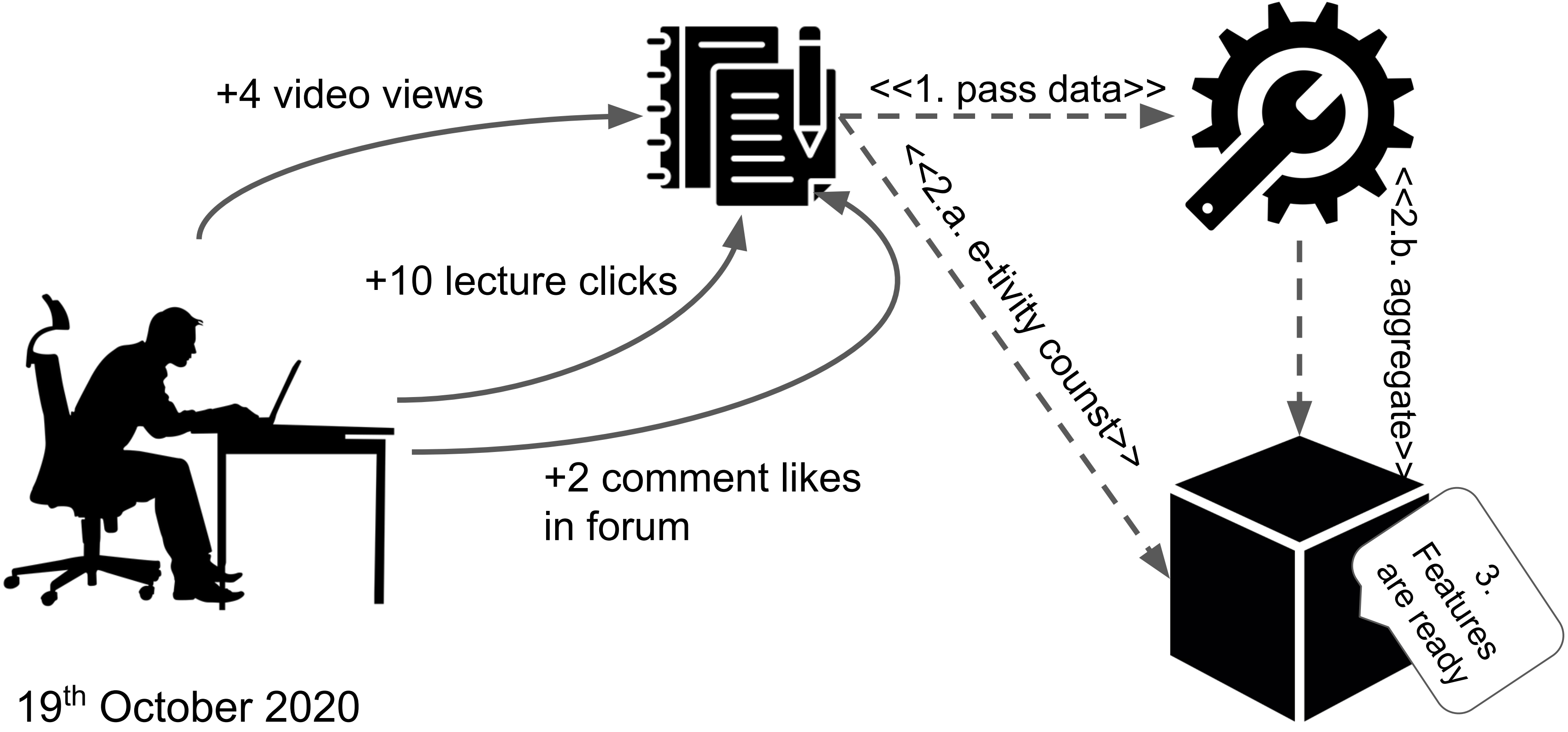
- Clicking resource e-tivities (e.g. page-view, video-view)

- Aggregate same type click e-tivities of each student

- **NO** forum thread discussions

- **NO** homework submission details

Clickstream-based schema



19th October 2020

Clickstream-based schema (e.g.)

- Weekly aggregation of page and video views [41]
- Lecture views and quiz answers for active/passive engagements [65]
- Clickstream data from video viewing patterns [59,62]

[41] Marius Kloft, Felix Stiehler, Zhilin Zheng, and Niels Pinkwart. 2014. Predicting MOOC dropout over weeks using machine learning methods. In Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs 60–65

[65] Arti Ramesh, Dan Goldwasser, Bert Huang, Hal Daume III, and Lise Getoor. 2014. Learning latent engagement patterns of students in online courses. In 28th AAAI Conference on Artificial Intelligence.

[59] Saurabh Nagrecha, John Z. Dillon, and Nitesh V. Chawla. 2017. MOOC dropout prediction: lessons learned from making pipelines interpretable. In Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee. 351–359.

[62] Jiezhong Qiu, Jie Tang, Tracy Xiao Liu, Jie Gong, Chenhui Zhang, Qian Zhang, and Yufei Xue. 2016. Modeling and predicting learning behavior in MOOCs. In Proceedings of the 9th ACM Int. Conf. on WSDM.. ACM, ACM, New York, NY, USA, 93–102.

Clickstream-based schema (e.g.)

- Homework submissions, grades, and user clicks [33]
- Lecture views, downloads, and quiz attempts [24]
- Behavioural patterns [50] and characteristics from weekly records [14]

[33] Jiazhen He, James Bailey, Benjamin I.P. Rubinstein, and Rui Zhang. 2015. Identifying at-risk students in massive open online courses. In 29th AAAI Conference on Artificial Intelligence.

[24] Mi Fei and Dit-Yan Yeung. 2015. Temporal models for predicting student dropout in massive open online courses. In 2015 IEEE ICDMW. IEEE, 256–263.

[50] Wentao Li, Min Gao, Hua Li, Qingyu Xiong, Junhao Wen, and Zhongfu Wu. 2016. Dropout prediction in MOOCs using behavior features and multi-view semi-supervised learning. In 2016 IJCNN, IEEE, 3130–3137.

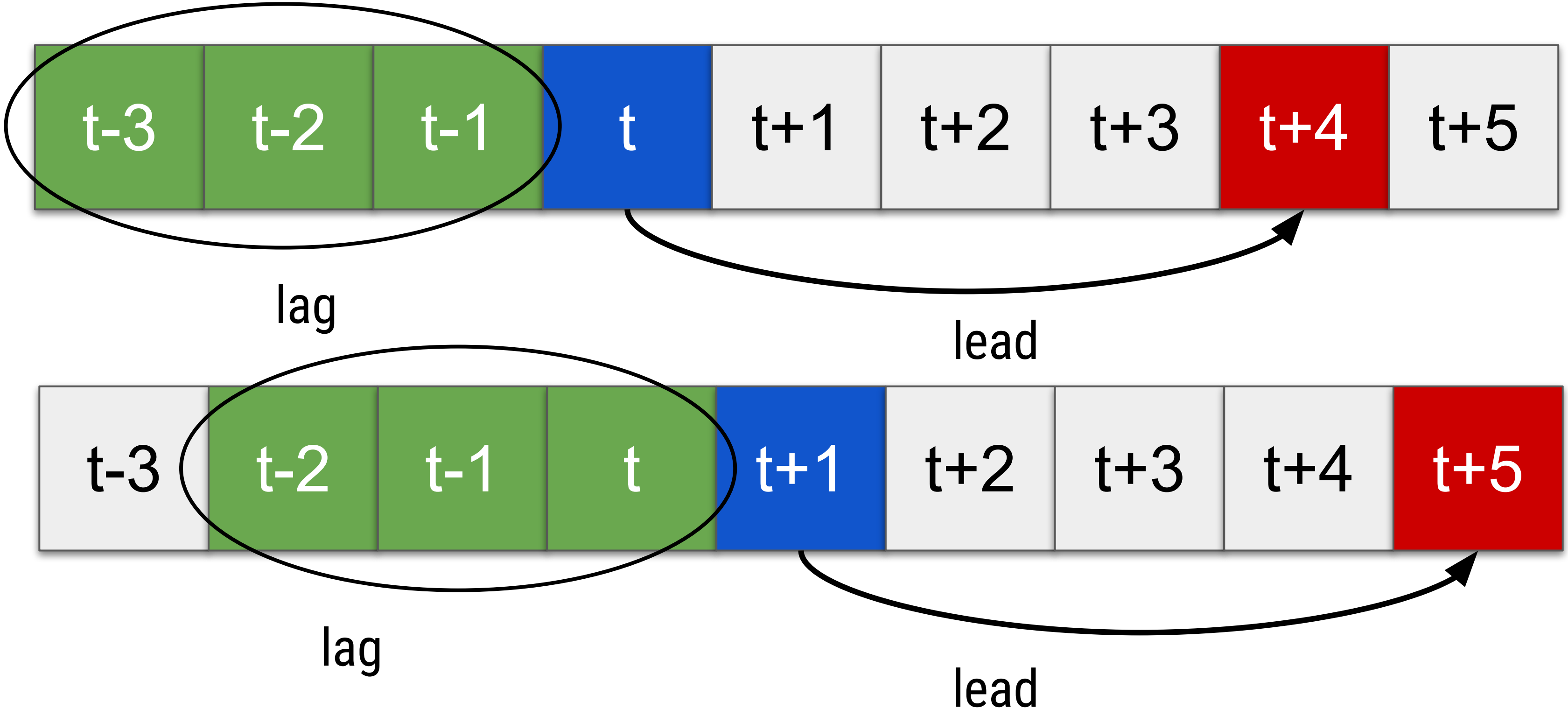
[14] Jing Chen, Jun Feng, Xia Sun, Nannan Wu, Zhengzheng Yang, and Sushing Chen. 2019. MOOC Dropout Prediction Using a Hybrid Algorithm Based on Decision Tree and Extreme Learning Machine. Mathematical Problems in Engineering 2019 (2019)

Clickstream-based schema (e.g. [72])

- Compute statistics at the end of a week
  - Average number of attempts on each assignment done by week i
- Introduction of **lag** and **lead**
- Use a portion (lag) of the history of a student and predict in the future (lead)

[72] Colin Taylor, Kalyan Veeramachaneni, and Una-May O'Reilly. 2014. Likely to stop? predicting stopout in massive open online courses.

Clickstream-based schema (e.g. [72])



[72] Colin Taylor, Kalyan Veeramachaneni, and Una-May O'Reilly. 2014. Likely to stop? predicting stopout in massive open online courses.

Forum-intervention-based schema

- Students discuss with their peers and tutors
- Thread initiatives, comments, and replies
- Use forum-derived data and NLP metrics in each course phase
  - Total number of responses
  - Text length/density
  - ...

Forum-intervention-based schema (e.g.)

- Linguistic characteristics and structural typology [65]

- Investigate forum views, thread initiations, posts & comments [24]

- Derived features and student homophily correlation [62]

[65] Arti Ramesh, Dan Goldwasser, Bert Huang, Hal Daume III, and Lise Getoor. 2014. Learning latent engagement patterns of students in online courses. In 28th AAAI Conference on Artificial Intelligence.

[24] Mi Fei and Dit-Yan Yeung. 2015. Temporal models for predicting student dropout in massive open online courses. In 2015 IEEE ICDMW. IEEE, 256–263.

[62] Jiezhong Qiu, Jie Tang, Tracy Xiao Liu, Jie Gong, Chenhui Zhang, Qian Zhang, and Yufei Xue. 2016. Modeling and predicting learning behavior in MOOCs. In Proceedings of the 9th ACM Int. Conf. on WSDM.. ACM, ACM, New York, NY, USA, 93–102.

Temporal networks

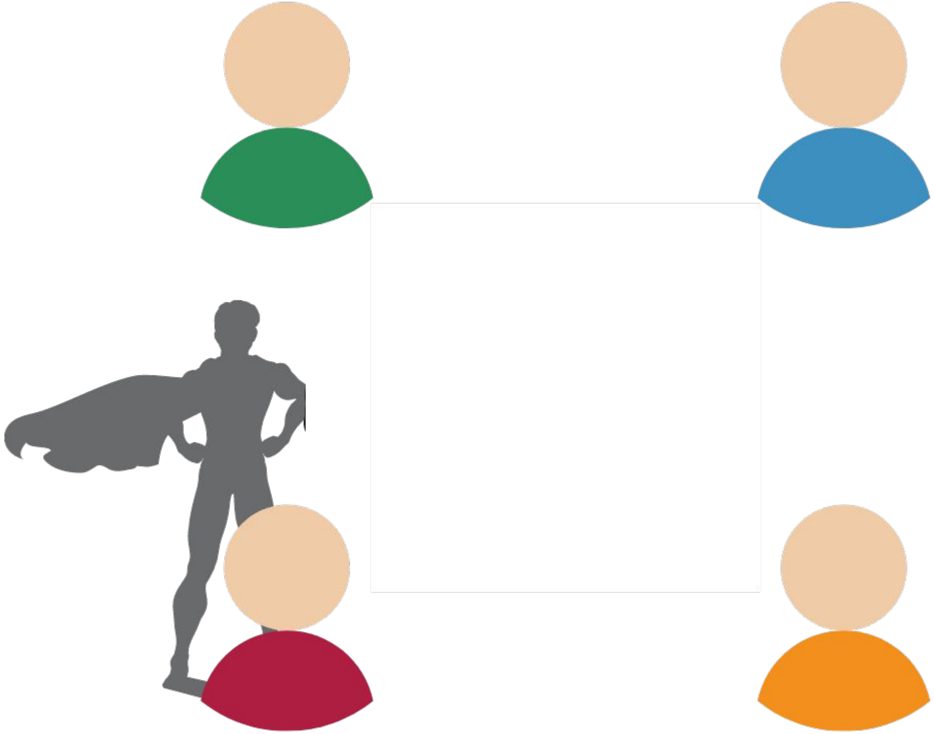
- Distinguish hugely **engaging** students from less determined students

- Commenting and replying to threads suggest a **lower** dropout probability

- Capture both **temporal & structural** properties

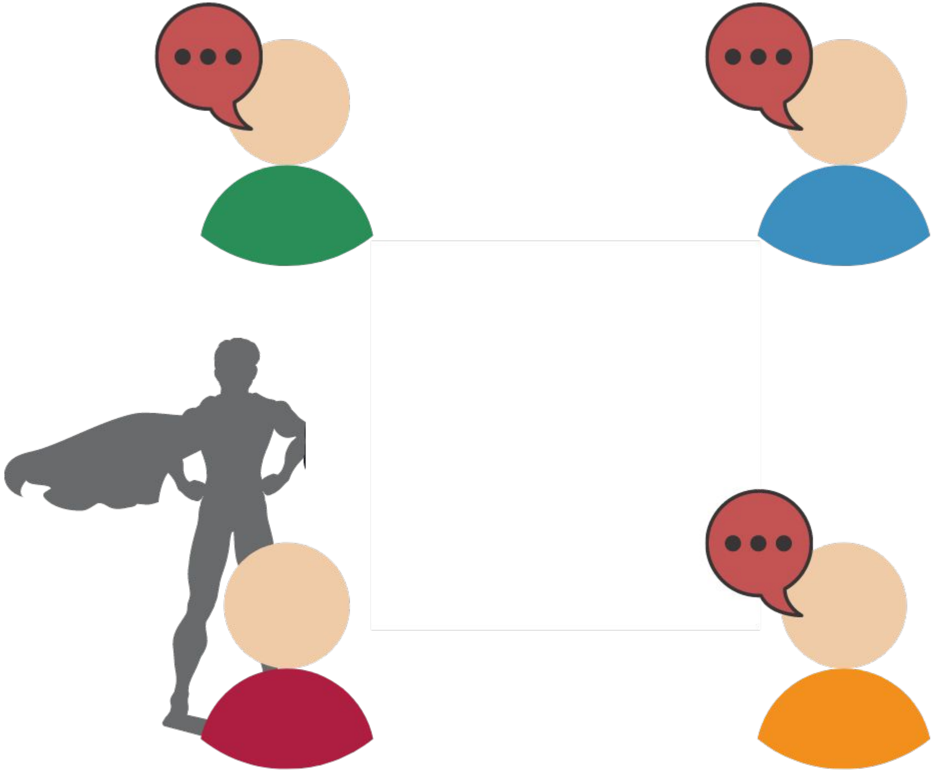
Temporal networks

- **Thread init** = the action of creating a new argument *θ* w/ [0,n] comments. The student that inits the thread is noted as *OP(θ)*



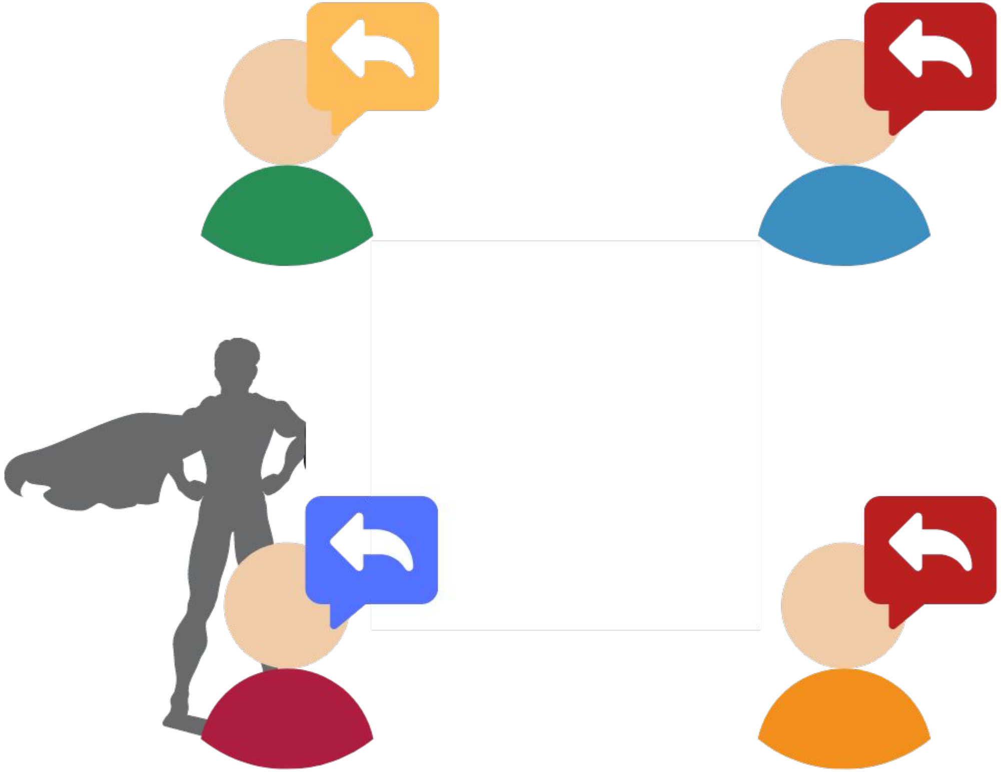
Temporal networks

- **Comment** = posts directly related to the thread w/ [0,n] replies



Temporal networks

- **Reply** = responses to comment messages w/ other nested replies if applicable



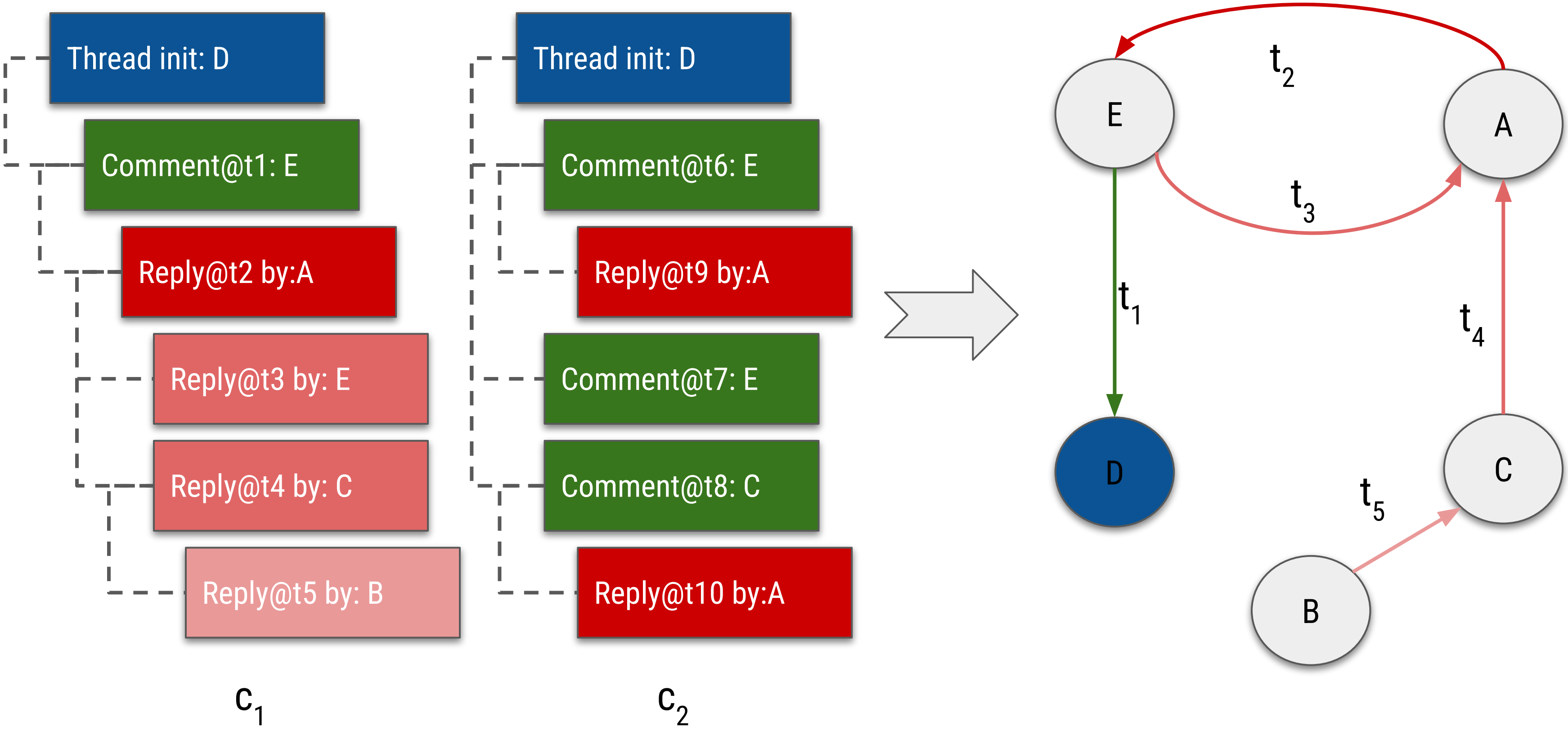
Temporal networks (formalisation [6,56])

- $\Theta_c = \{\theta^1, \theta^2, \ldots, . \theta^x\}$
- Forum-based network for $\theta^r \in \Theta_c$ in [t_b,t_f] = labelled multidigraph $\mathcal{M}^{\theta^r}_{[t_b,t_f]} = (V, A, \ell_A)$
- A is the set of arcs (s’,s”)
  - s’ comments thread θ^r that s” has generated
  - s’ replies to a comment of s” in thread θ^r
- $\ell_A : A \rightarrow \mathcal{T}$ is a function that maps the edges (s’,s”) to timestamps

[6] V.K. Balakrishnan. 1997. Schaum’s Outline of Graph Theory: Including Hundreds of Solved Problems . McGraw Hill Professional, New York, NY, USA.

[56] Othon Michail. 2016. An introduction to temporal graphs: An algorithmic perspective. Internet Mathematics 12, 4 (2016), 239–280

Temporal networks (e.g.)



Thread init: D

Comment@t1: E

Reply@t2 by:A

Reply@t3 by: E

Reply@t4 by: C

Reply@t5 by: B

c_1

Thread init: D

Comment@t6: E

Reply@t9 by:A

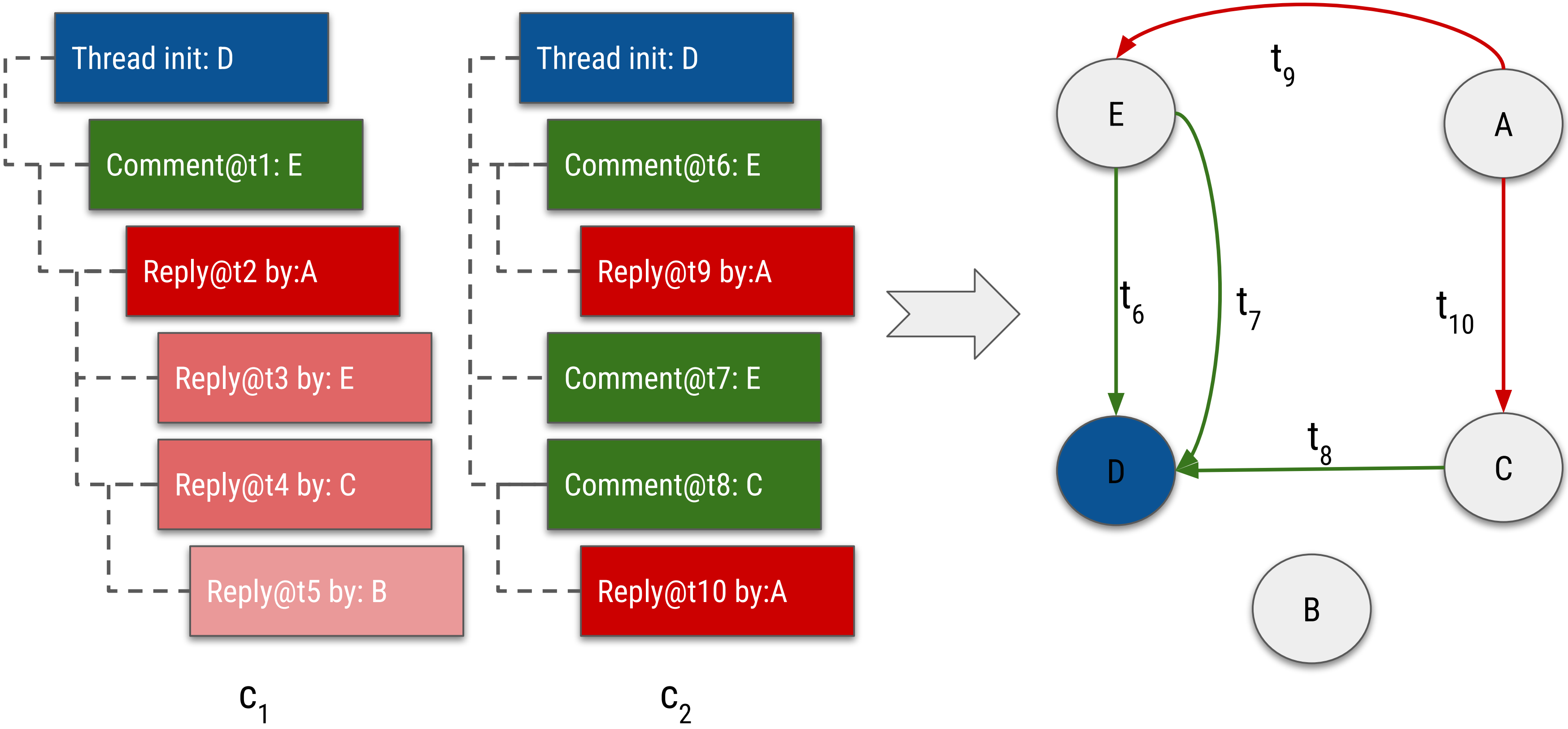
Comment@t7: E

Comment@t8: C

Reply@t10 by:A

c_2

Temporal networks (e.g.)



Thread init: D

Comment@t1: E

Reply@t2 by:A

Reply@t3 by: E

Reply@t4 by: C

Reply@t5 by: B

c_1

Thread init: D

Comment@t6: E

Reply@t9 by:A

Comment@t7: E

Comment@t8: C

Reply@t10 by:A

c_2

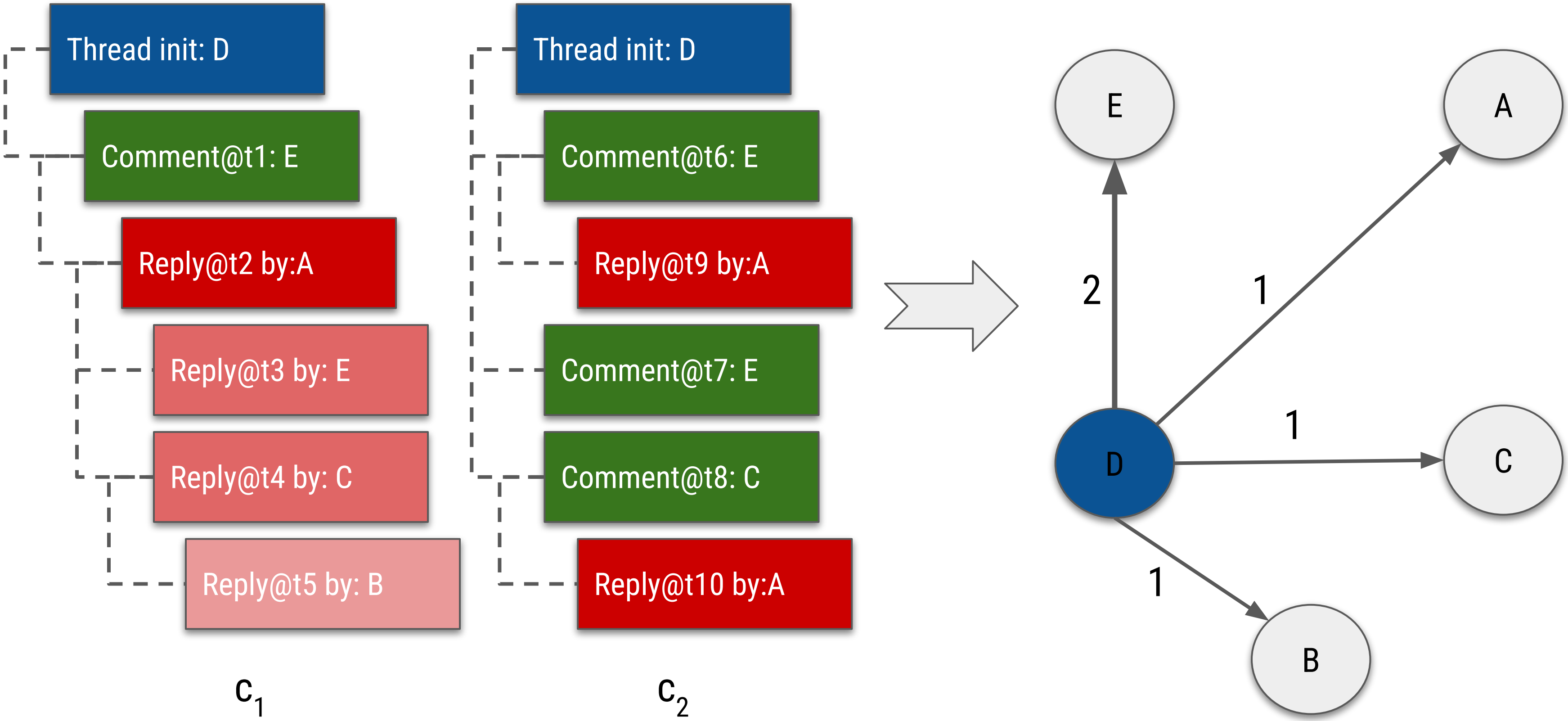
Thread initiator w/ outgoing edges (e.g. [77])

- Thread initiators have outward links to all students participating in the discussion

- **Transpose** our multidigraph in each course phase
  - No parallel connections
  - Label multiple connections with the number of interactions between the end-points

[77] Diyi Yang, Tanmay Sinha, David Adamson, and Carolyn Penstein Rosé. 2013. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In Proceedings of the 2013 NIPS Data-driven education workshop. Vol. 11. Curran Associates, Inc., USA, 14.

Thread initiator w/ outgoing edges (e.g. [77])



Thread init: D

Comment@t1: E

Reply@t2 by:A

Reply@t3 by: E

Reply@t4 by: C

Reply@t5 by: B

c_1

Thread init: D

Comment@t6: E

Reply@t9 by:A

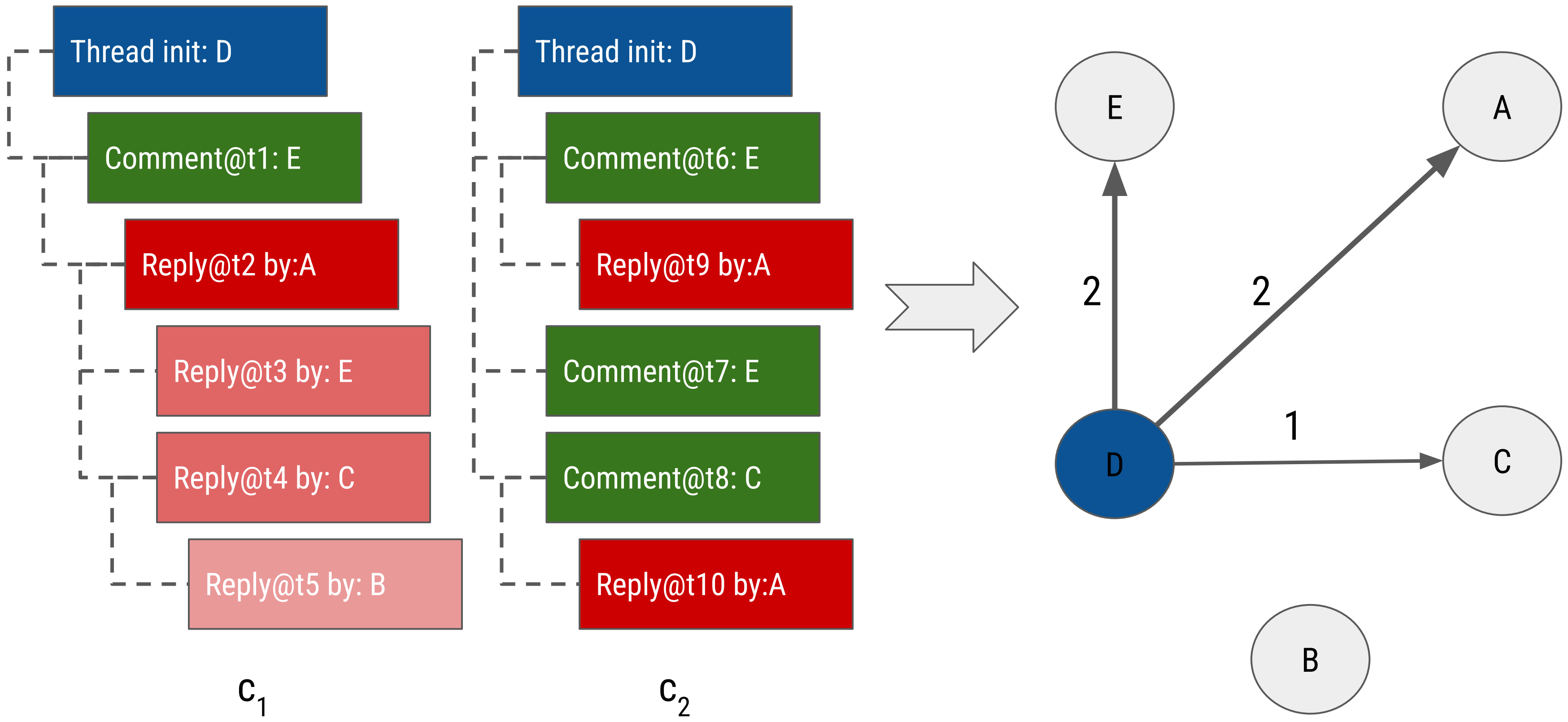
Comment@t7: E

Comment@t8: C

Reply@t10 by:A

c_2

Thread initiator w/ outgoing edges (e.g. [77])



Thread init: D

Comment@t1: E

Reply@t2 by:A

Reply@t3 by: E

Reply@t4 by: C

Reply@t5 by: B

c_1

Thread init: D

Comment@t6: E

Reply@t9 by:A

Comment@t7: E

Comment@t8: C

Reply@t10 by:A

c_2

Newcomers respond to all (e.g. [11,30])

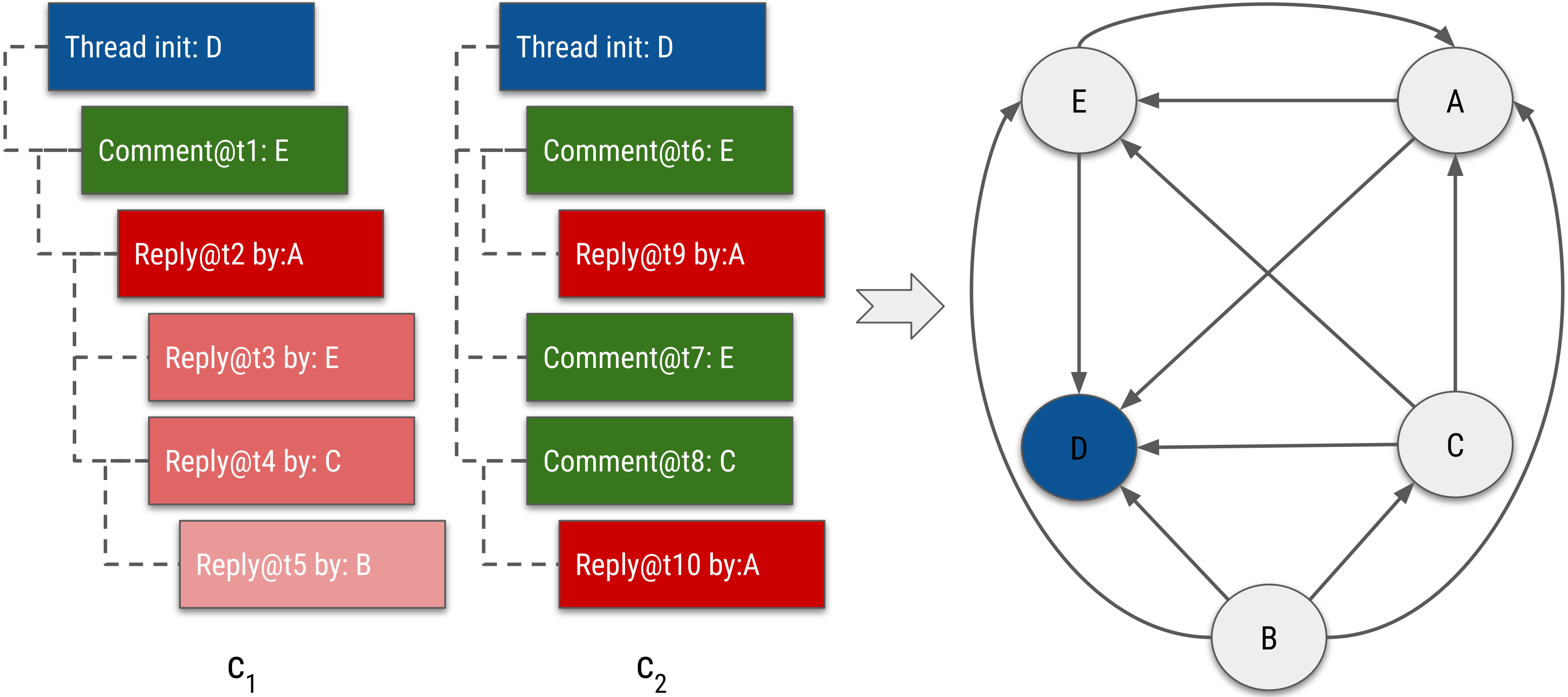
- **Assumption**: every new participant has read the previous comments/replies in a thread = **respond to everyone**

- **Transpose** our multidigraph in each course phase
  - No parallel connections
  - No labels for multiple connections between the same endpoints

[11] Rebecca Brown, Collin Lynch, Yuan Wang, Michael Eagle, Jennifer Albert, Tiffany Barnes, Ryan Shaun Baker, Yoav Bergner, and Danielle S. McNamara. 2015. Communities of Performance & Communities of Preference. In CEUR Workshop Proceedings, Vol. 1446. CEUR-WS, USA

[30] Niki Gitinabard, Farzaneh Khoshnevisan, Collin F. Lynch, and Elle Yuan Wang. 2018. Your Actions or Your Associates? Predicting Certification and Dropout in MOOCs with Behavioral and Social Features.

Newcomers respond to all (e.g. [11,30])



Thread init: D

Comment@t1: E

Reply@t2 by:A

Reply@t3 by: E

Reply@t4 by: C

Reply@t5 by: B

c_1

Thread init: D

Comment@t6: E

Reply@t9 by:A

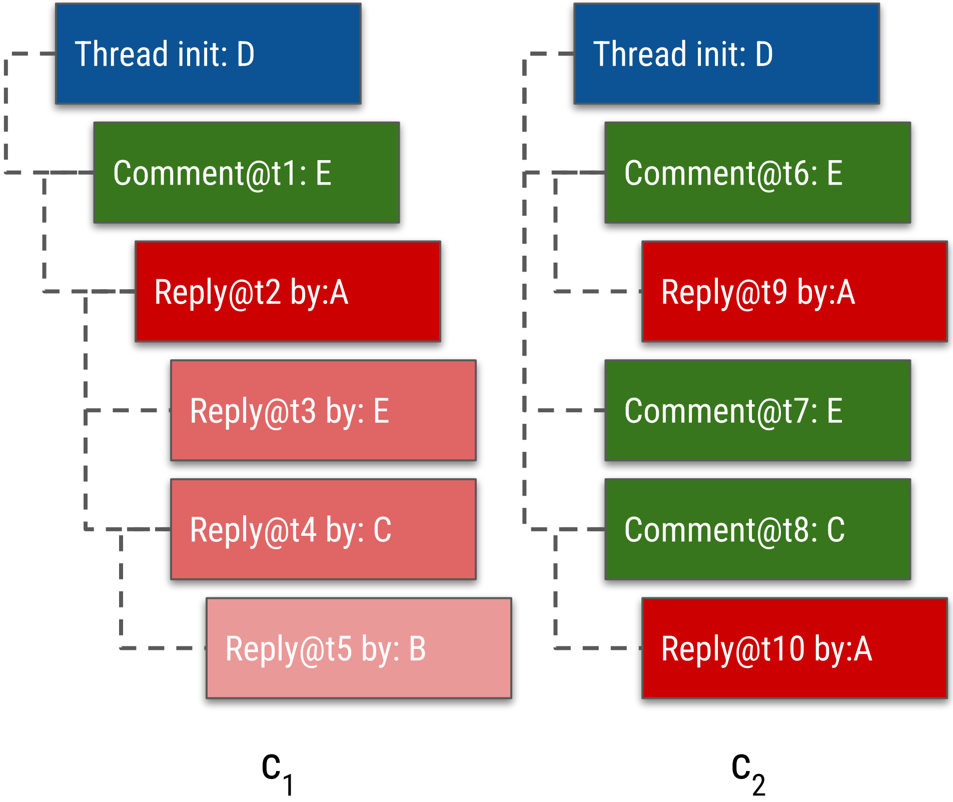
Comment@t7: E

Comment@t8: C

Reply@t10 by:A

c_2

Newcomers respond to all (e.g. [11,30])



Thread init: D

Comment@t1: E

Reply@t2 by:A

Reply@t3 by: E

Reply@t4 by: C

Reply@t5 by: B

c_1

Thread init: D

Comment@t6: E

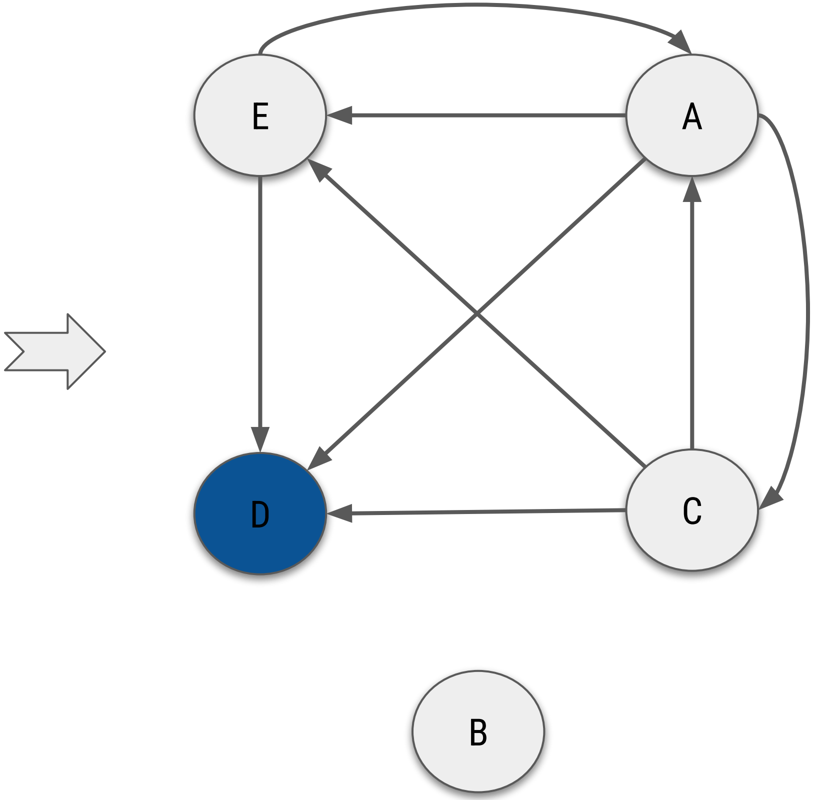
Reply@t9 by:A

Comment@t7: E

Comment@t8: C

Reply@t10 by:A

c_2

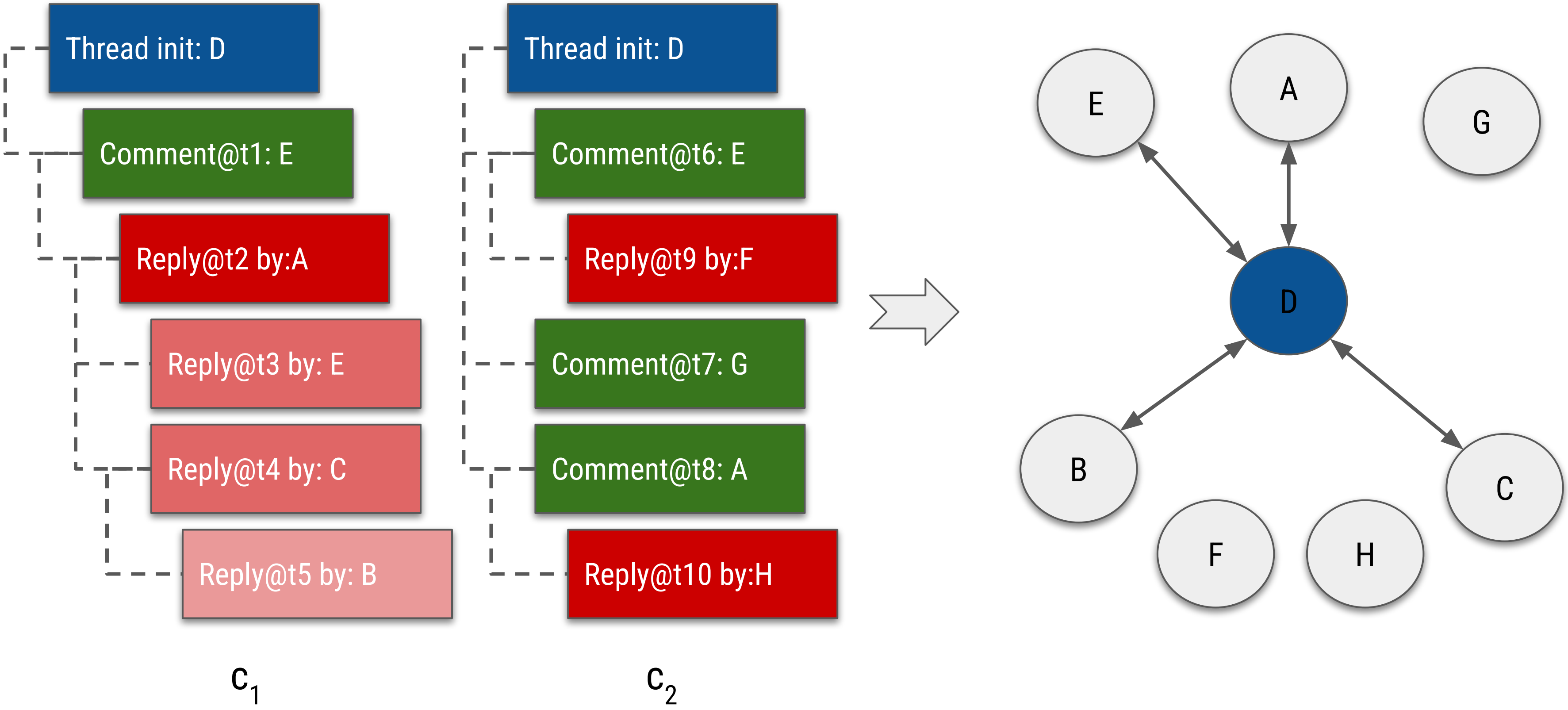


Temporal star network (e.g. [30,80])

- **Thread-tracing** networks
- Comments and replies are the same
- Connect all the students with OP(θ)
- **Transform** our multidigraph in each course phase
  - Bidirectional edges between OP(θ) and all participants
  - Thread starters are at the centre of a star network

[80] Mengxiao Zhu, Yoav Bergner, Yan Zhan, Ryan Baker, Yuan Wang, and Luc Paquette. 2016. Longitudinal engagement, performance, and social connectivity: a MOOC case study using exponential random graph models. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge. ACM, ACM, New York, NY, USA, 223–230

Temporal star network (e.g. [30,80])



Thread init: D

Comment@t1: E

Reply@t2 by:A

Reply@t3 by: E

Reply@t4 by: C

Reply@t5 by: B

c_1

Thread init: D

Comment@t6: E

Reply@t9 by:F

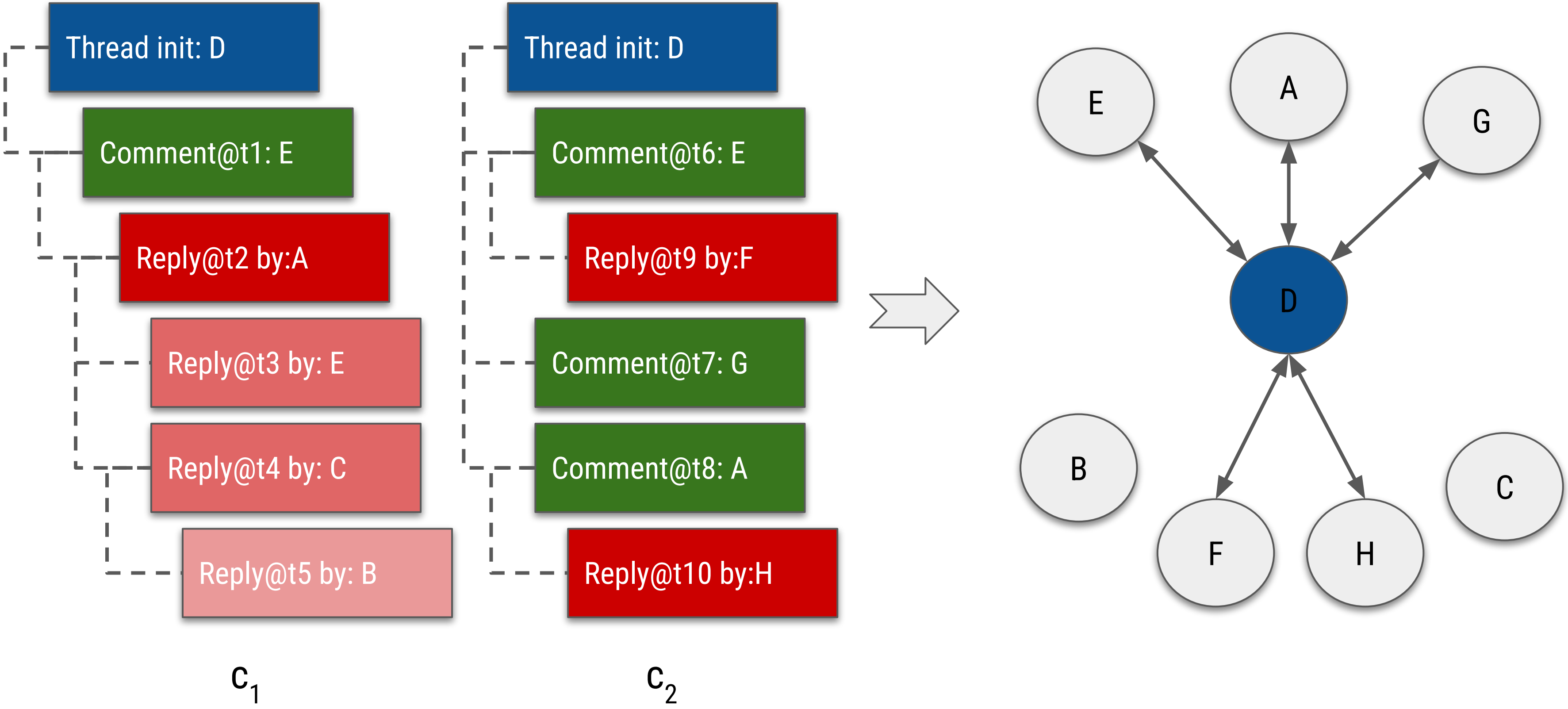
Comment@t7: G

Comment@t8: A

Reply@t10 by:H

c_2

Temporal star network (e.g. [30,80])

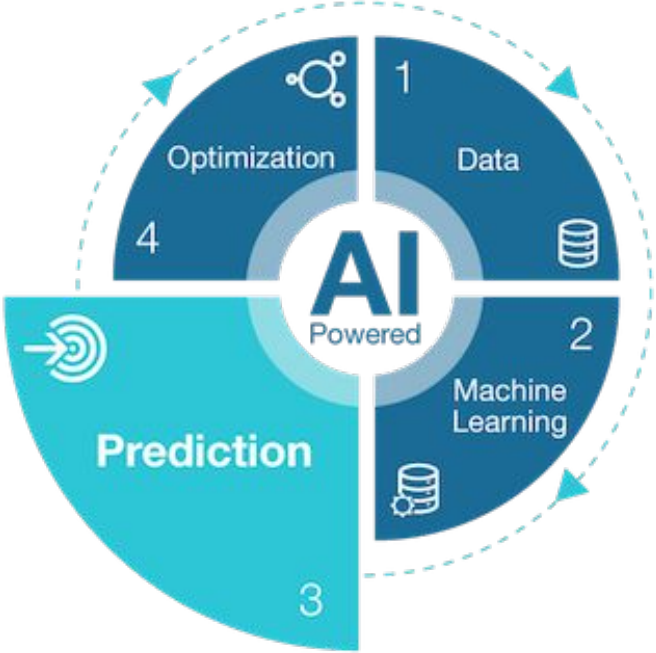


Part Three

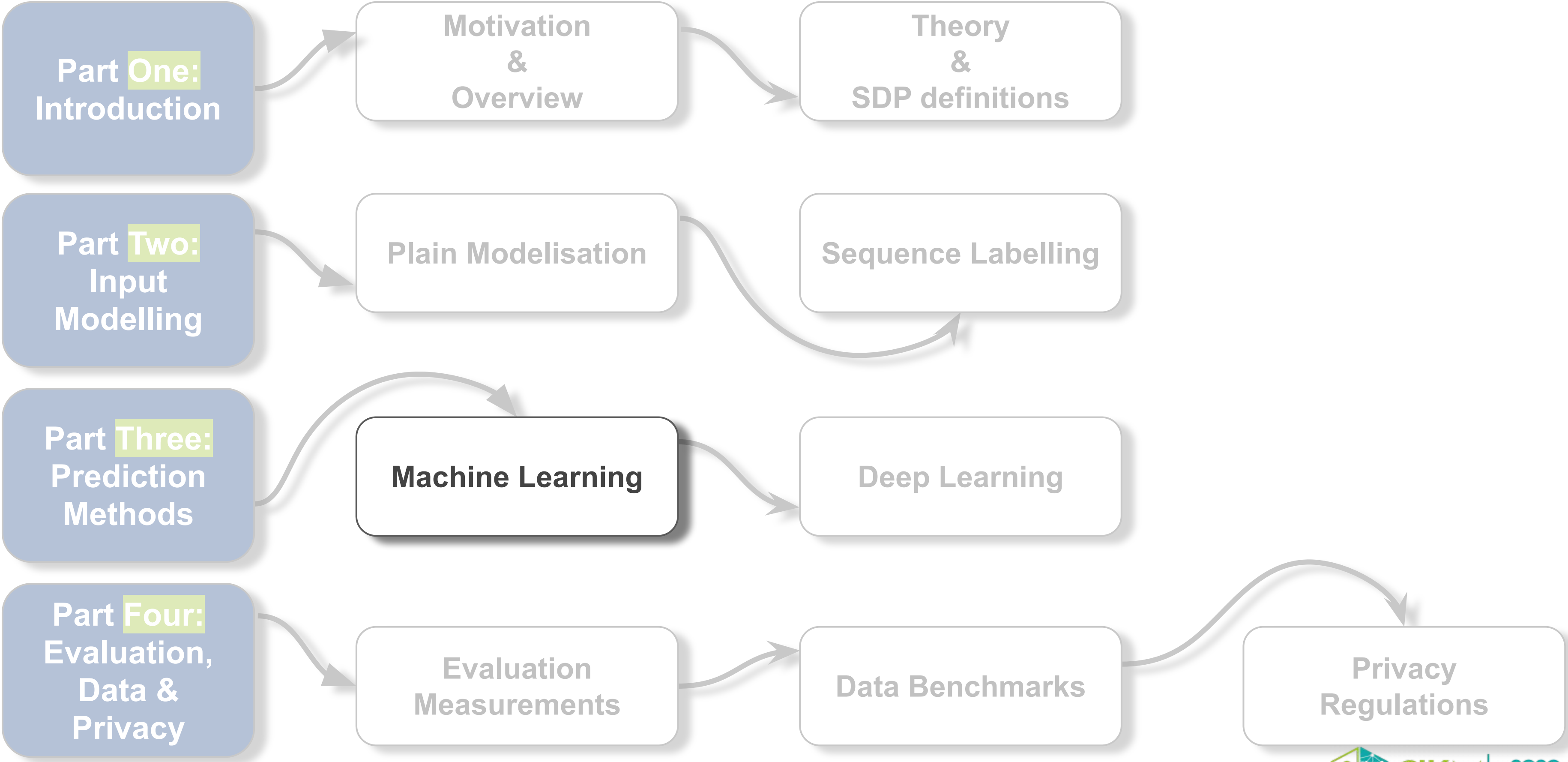
Prediction Methods

Prediction methods in SDP

- **Classic learning methods**
  - Typical machine learning algorithms
  - Widely supported by programmatic frameworks
  - **Mostly plain modelisation**
- **Deep learning strategies**
  - Complex strategies based on recurrent and convolutional neural networks
  - **Avoid manual feature engineering**



Roadmap



Part One: Introduction

Motivation & Overview

Theory & SDP definitions

Part Two: Input Modelling

Plain Modelisation

Sequence Labelling

Part Three: Prediction Methods

Machine Learning

Deep Learning

Part Four: Evaluation, Data & Privacy

Evaluation Measurements

Data Benchmarks

Privacy Regulations

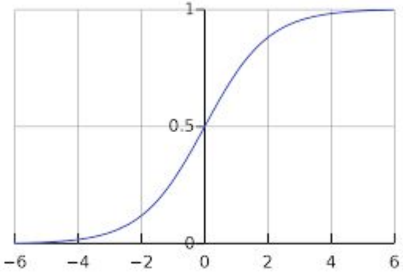
Classic learning methods

- **Off-the-shelf:**
  - Logistic Regression
  - Naive Bayes Classifiers
  - Support-Vector Machine
  - Decision Tree
- **Ensembles:**
  - Random Forest
  - AdaBoost
- **Feed-Forward Neural Network**

Logistic Regression

Logistic Regression is a supervised probabilistic binary classifier that attempts to learn a decision function by estimating the classification probability of a given vector of features' values.

$$y* = P(y = 1|x) = \sigma(w^T x + b)$$

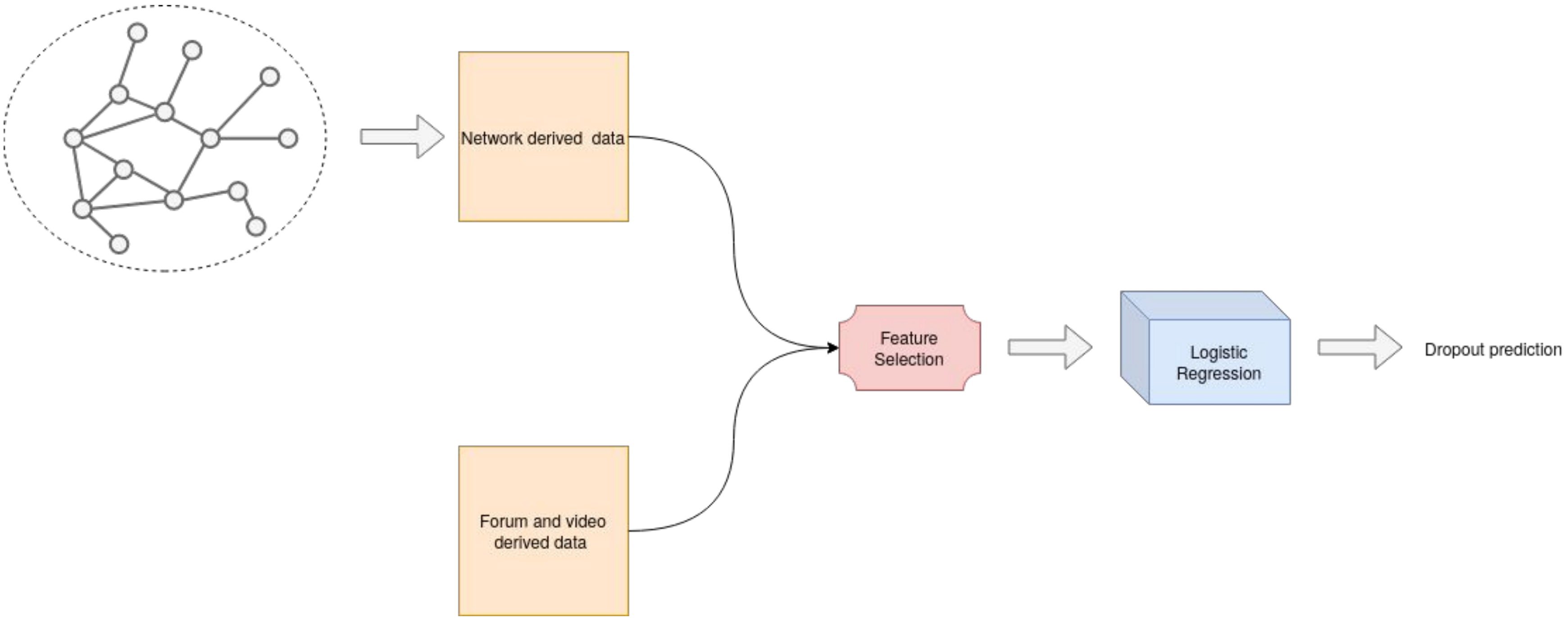


- Adjust weights to reduce the error

$$f(z) = \frac{1}{1+e^{-z}}$$

[30] Niki Gitinabard, Farzaneh Khoshnevisan, Collin F. Lynch, and Elle Yuan Wang. 2018. Your actions or your associates? Predicting certification and dropout in MOOCs with behavioral and social features.
[33] Jiazhen He, James Bailey, Benjamin I. P. Rubinstein, and Rui Zhang. 2015. Identifying at-risk students in massive open online courses. In Proceedings of the 29th AAAI Conference on Artificial Intelligence.
[66] Carly Robinson, Michael Yeomans, Justin Reich, Chris Hulleman, and Hunter Gehlbach. 2016. Forecasting student achievement in MOOCs with natural language processing. In Proceedings of the 6th International Conference on Learning Analytics & Knowledge. ACM, New York, NY, 383–387.
[72] Colin Taylor, Kalyan Veeramachaneni, and Una-May O'Reilly. 2014. Likely to stop? Predicting stopout in massive open online courses.

Logistic Regression (e.g. [30])



[30] Niki Gitinabard, Farzaneh Khoshnevisan, Collin F. Lynch, and Elle Yuan Wang. 2018. Your actions or your associates? Predicting certification and dropout in MOOCs with behavioral and social features.

Naive Bayes Classifiers

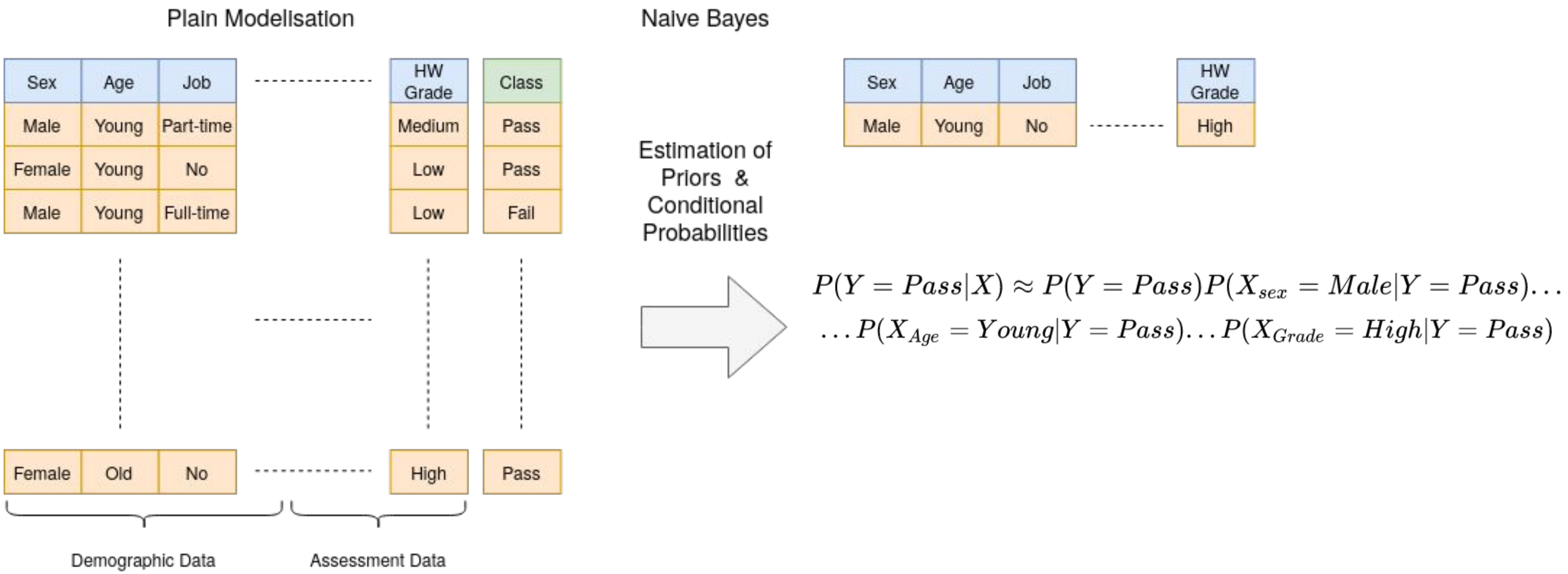
Naive Bayes classifiers are a set of supervised probabilistic classifiers based on Bayes' theorem with the "naive" assumption.

$$y^* \approx argmax_{y_i} = P(Y = y_i|X = x_j) = P(Y = y_i)\prod_{j=1}^{d} P(X = x_j|Y = y_i)$$

- **Naive assumption**: Each feature contributes independently to the outcome
- Naive Bayes estimates the probabilities by Maximum Likelihood

[29] Elena Gaudioso, Miguel Montero, and Felix Hernandez-Del-Olmo. 2012. Supporting teachers in adaptive educational systems through predictive models: A proof of concept. Exp. Syst. Applic. 39, 1 (2012), 621–625.
[44] Sotiris Kotsiantis, Christos Pierrakeas, and Panagiotis Pintelas. 2003. Preventing student dropout in distance learning using machine learning techniques. In Proceedings of the International Conference on Knowledge-based and Intelligent Information and Engineering Systems. Springer, New York, NY, 267–274.
[45] Sotiris Kotsiantis, Christos Pierrakeas, Ioannis Zaharakis, and Panagiotis Pintelas. 2003. Efficiency of Machine Learning Techniques in Predicting Students Performance in Distance Learning Systems. University of Patras Press, 297–306.
[50] Wentao Li, Min Gao, Hua Li, Qingyu Xiong, Junhao Wen, and Zhongfu Wu. 2016. Dropout prediction in MOOCs using behavior features and multi view semi-supervised learning. In Proceedings of the International Joint Conference on Neural Networks (IJCNN'16). IEEE, 3130–3137.
[54] Laci Mary Barbosa Manhães, Sérgio Manuel Serra da Cruz, and Geraldo Zimbrão. 2014. WAVE: An architecture for predicting dropout in undergraduate courses using EDM. In Proceedings of the 29th ACM Symposium on Applied Computing. ACM, New York, NY, 243–247.

Naive Bayes Classifiers (e.g. [45])



Plain Modelisation

Sex	Age	Job	...	HW Grade	Class
Male	Young	Part-time		Medium	Pass
Female	Young	No		Low	Pass
Male	Young	Full-time		Low	Fail
...	...	...	...	...	...
Female	Old	No	...	High	Pass

Demographic Data

Assessment Data

Estimation of Priors & Conditional Probabilities

Naive Bayes

Sex	Age	Job	...	HW Grade
Male	Young	No	...	High

$$P(Y = Pass|X) \approx P(Y = Pass)P(X_{sex} = Male|Y = Pass)\dots$$

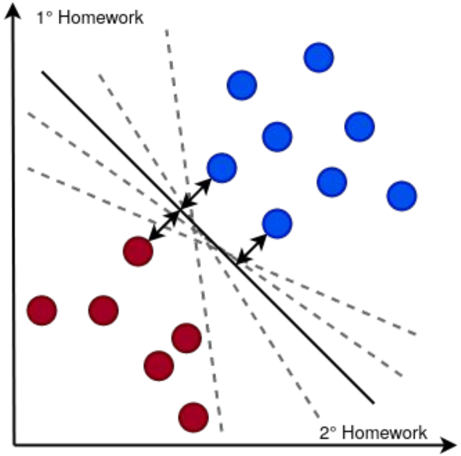
$$\dots P(X_{Age} = Young|Y = Pass)\dots P(X_{Grade} = High|Y = Pass)$$

[45] Sotiris Kotsiantis, Christos Pierrakeas, Ioannis Zaharakis, and Panagiotis Pintelas. 2003. Efficiency of Machine Learning Techniques in Predicting Students Performance in Distance Learning Systems. University of Patras Press, 297–306.

Support-Vector Machine

Support-Vector Machine (SVM) is a supervised algorithm that attempts to separate classes of data in the feature space using a hyperplane.

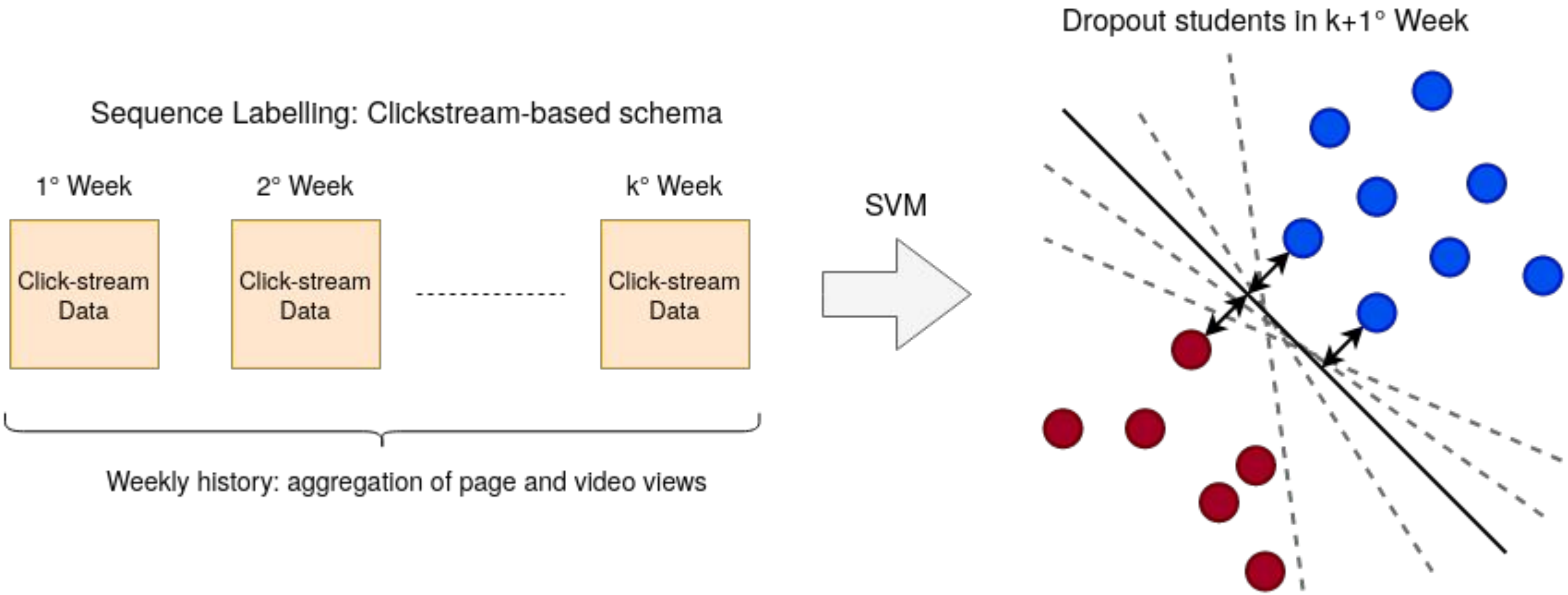
- Handle nonlinear feature space



[3] Bussaba Amnueypornsakul, Suma Bhat, and Phakpoom Chinprutthiwong. 2014. Predicting attrition along the way: The UIUC model. In Proceedings of the EMNLP Workshop on Analysis of Large Scale Social Interaction in MOOCs. Association for Computational Linguistics, 55–59.

[41] Marius Kloft, Felix Stiehler, Zhilin Zheng, and Niels Pinkwart. 2014. Predicting MOOC dropout over weeks using machine learning methods. In Proceedings of the EMNLP Workshop on Analysis of Large Scale Social Interaction in MOOCs. 60–65.

Support-Vector Machine (e.g. [41])

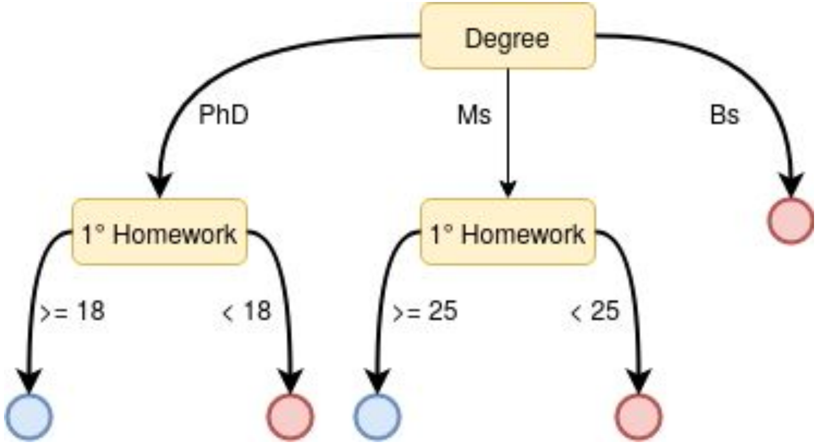


Dropout students in k+1° Week
Sequence Labelling: Clickstream-based schema
1° Week
2° Week
k° Week
Click-stream Data
Click-stream Data
Click-stream Data
SVM
Weekly history: aggregation of page and video views

[41] Marius Kloft, Felix Stiehler, Zhilin Zheng, and Niels Pinkwart. 2014. Predicting MOOC dropout over weeks using machine learning methods. In Proceedings of the EMNLP Workshop on Analysis of Large Scale Social Interaction in MOOCs. 60–65.

Decision Tree

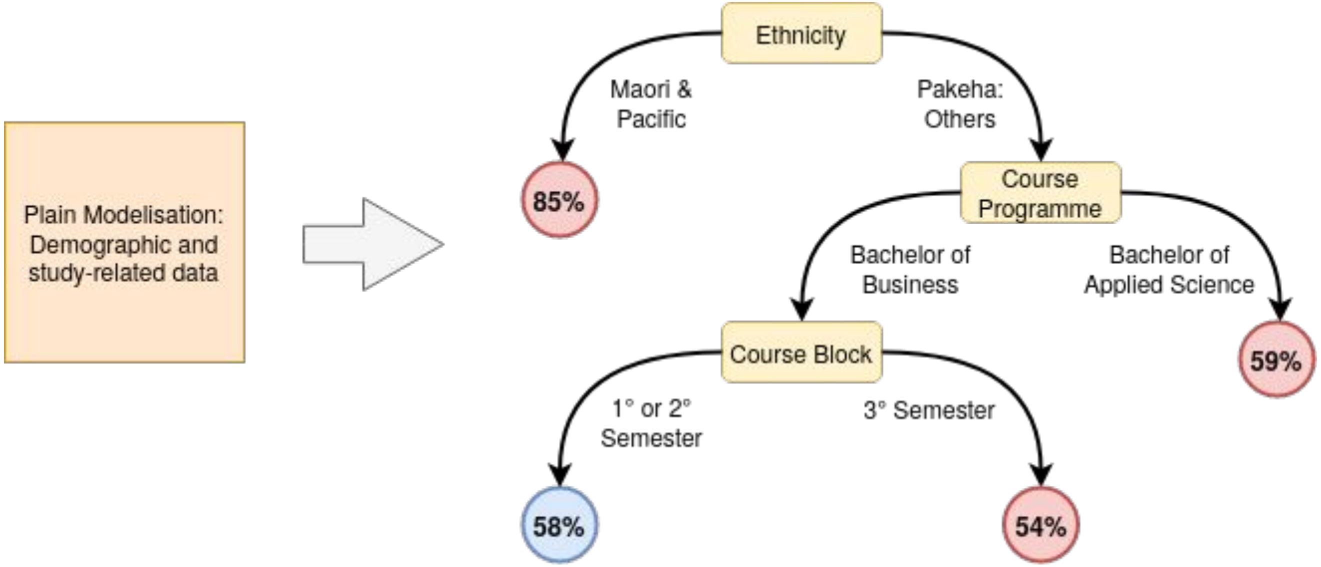
Decision Tree (dtree) is a rule-based supervised classifier that infers decision rules in a tree-like structure from the input data features.



[1] Qasem A. Al-Radaideh, Emad M. Al-Shawakfa, and Mustafa I. Al-Najjar. 2006. Mining student data using decision trees. In Proceedings of the International Arab Conference on Information Technology (ACIT'2006). 1–5.
[20] Gerben W. Dekker, Mykola Pechenizkiy, and Jan M. Vleeshouwer. 2009. Predicting students drop out: A case study. In Proceedings of the International Conference on Educational Data Mining (EDM'09).
[42] Georgios Kostopoulos, Sotiris Kotsiantis, and Panagiotis Pintelas. 2015. Estimating student dropout in distance higher education using semi-supervised techniques. In Proceedings of the 19th Panhellenic Conference on Informatics. ACM, New York, NY, 38–43.
[46] Zlatko J. Kovačić. 2010. Early prediction of student success: Mining student enrollment data. In Proceedings of the Informing Science & IT Education Conference. Citeseer.
[59] Saurabh Nagrecha, John Z. Dillon, and Nitesh V. Chawla. 2017. MOOC dropout prediction: Lessons learned from making pipelines interpretable. In Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, 351–359.

Decision Tree (e.g. [46])

Decision Tree for SDP in the Information Systems course at the Open Polytechnic of New Zealand



[46] Zlatko J. Kovačić. 2010. Early prediction of student success: Mining student enrollment data. In Proceedings of the Informing Science & IT Education Conference. Citeseer.

Ensembles

Principle: combine the outputs of several models to reduces the risk of making an incorrect prediction.

1. **Learning** phase: Multiple models are trained
2. **Aggregation** phase: The final prediction is the result of the aggregation of the outcomes of the trained models

- Resistant to overfitting

Ensembles: Random Forest

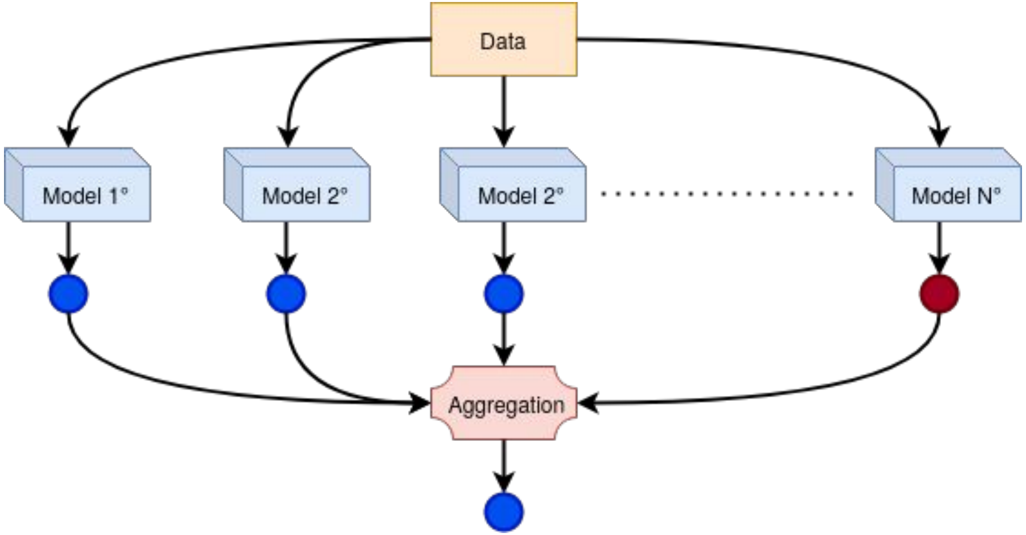
- **Learning phase**: Every d-tree is trained independently on a random subset of the input features

- **Aggregation phase**: The final outcome is the most predicted by the d-trees in the forest (i.e. majority voting)

[31] Cameron C. Gray and Dave Perkins. 2019. Utilizing early engagement and machine learning to predict student outcomes. Comput. Educ. 131 (2019), 22–32.
[32] Liu Haiyang, Zhihai Wang, Phillip Benachour, and Philip Tubman. 2018. A time series classification method for behaviour-based dropout prediction. In Proceedings of the IEEE 18th International Conference on Advanced Learning Technologies (ICALT'18). IEEE, 191–195.

Ensembles: Random Forests (e.g. [31],[32])

- In **[31]**, authors feed derived data to Random Forests to determine drop-out cases
- In **[32]**, authors use Time-Series Forests to identify the course period most affecting the learning process of a student.



[31] Cameron C. Gray and Dave Perkins. 2019. Utilizing early engagement and machine learning to predict student outcomes. Comput. Educ. 131 (2019), 22–32.
[32] Liu Haiyang, Zhihai Wang, Phillip Benachour, and Philip Tubman. 2018. A time series classification method for behaviour-based dropout prediction. In Proceedings of the IEEE 18th International Conference on Advanced Learning Technologies (ICALT'18). IEEE, 191–195.

Ensembles: AdaBoost

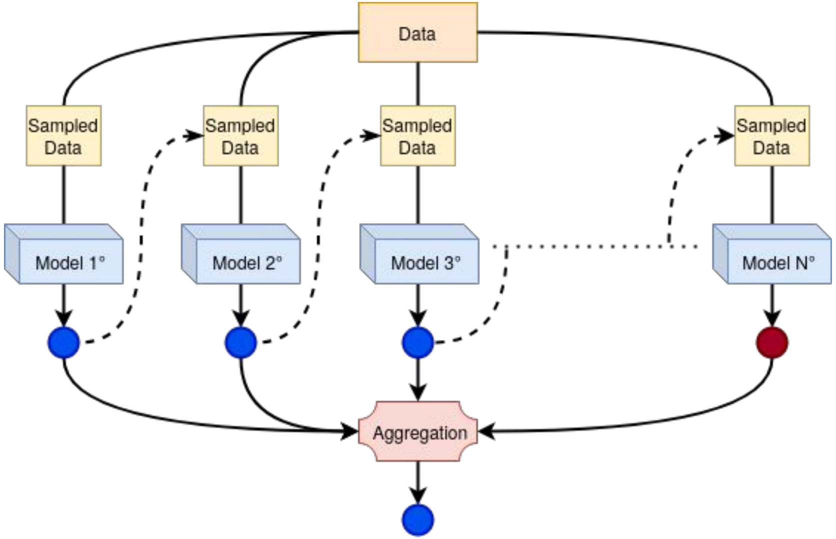
- **Learning phase**: Models are trained iteratively based on the performance of the previous model

- **Aggregation phase**: AdaBoost weights the contribution of each model to the final outcome (e.i. weighted voting)

[8] Johannes Berens, Kerstin Schneider, Simon Görtz, Simon Oster, and Julian Burghoff. 2019. Early detection of students at risk - Predicting student dropouts using administrative student data from German Universities and machine learning methods. Journal of Educational Data Mining 11, 3 (2019), 1–41.
[37] Ya-Han Hu, Chia-Lun Lo, and Sheng-Pao Shih. 2014. Developing early warning systems to predict students' online learning performance. Comput. Human Behav. 36 (2014), 469–478.

Ensembles: AdaBoost (e.g. [8], [37])

- In **[8]**, AdaBoost combines the prediction powers of linear regression, neural networks, and random forests to distinguish at-risk students and persisters.
- In **[37],** AdaBoost use decision trees to predict successful or failing students for a course

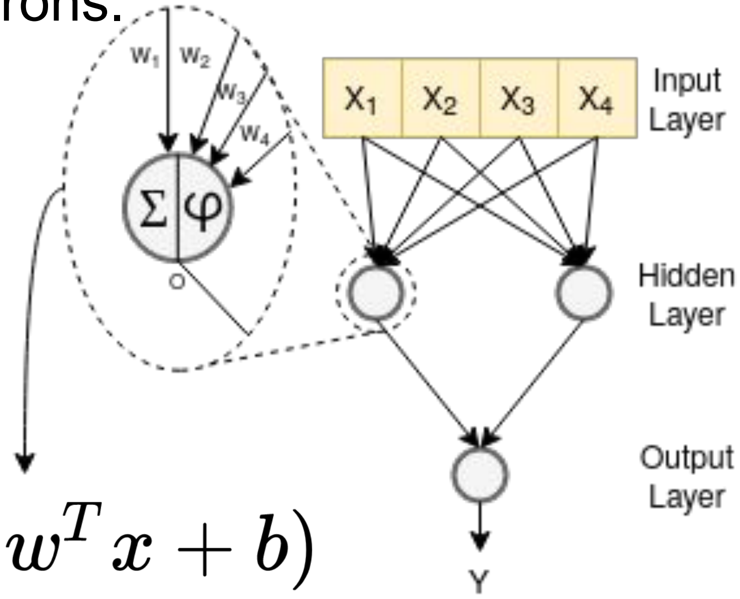


[8] Johannes Berens, Kerstin Schneider, Simon Görtz, Simon Oster, and Julian Burghoff. 2019. Early detection of students at risk - Predicting student dropouts using administrative student data from German Universities and machine learning methods. Journal of Educational Data Mining 11, 3 (2019), 1–41.
[37] Ya-Han Hu, Chia-Lun Lo, and Sheng-Pao Shih. 2014. Developing early warning systems to predict students' online learning performance. Comput. Human Behav. 36 (2014), 469–478.

Feed-Forward Neural Network

A Feed-Forward neural network (FFNN) is a layered graph architecture made up of computational units called neurons.

- Information is forwarded through the layers
- Error is back-propagated to adjust the weights

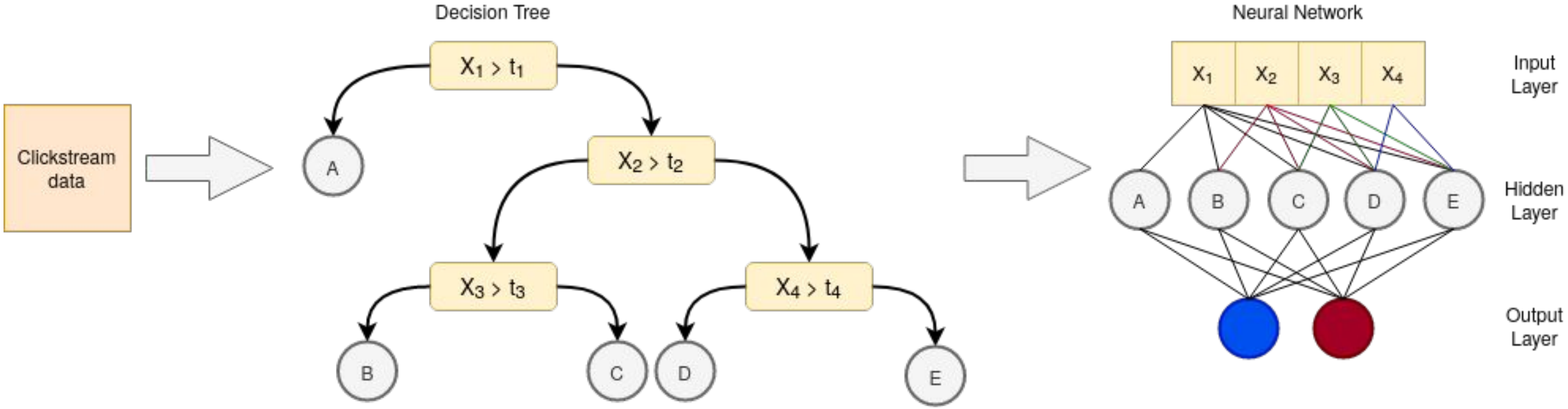


$$o = \varphi(w^T x + b)$$

[14] Jing Chen, Jun Feng, Xia Sun, Nannan Wu, Zhengzheng Yang, and Sushing Chen. 2019. MOOC dropout prediction using a hybrid algorithm based on decision tree and extreme learning machine. Math. Probl. Eng. 2019 (2019).

Feed-Forward Neural Network (e.g. [14])

An Extreme Learning Machine (ELM) is a FFNN with a different learning algorithm that does not need to tune the weights of hidden nodes. ELM need only one iteration to train.

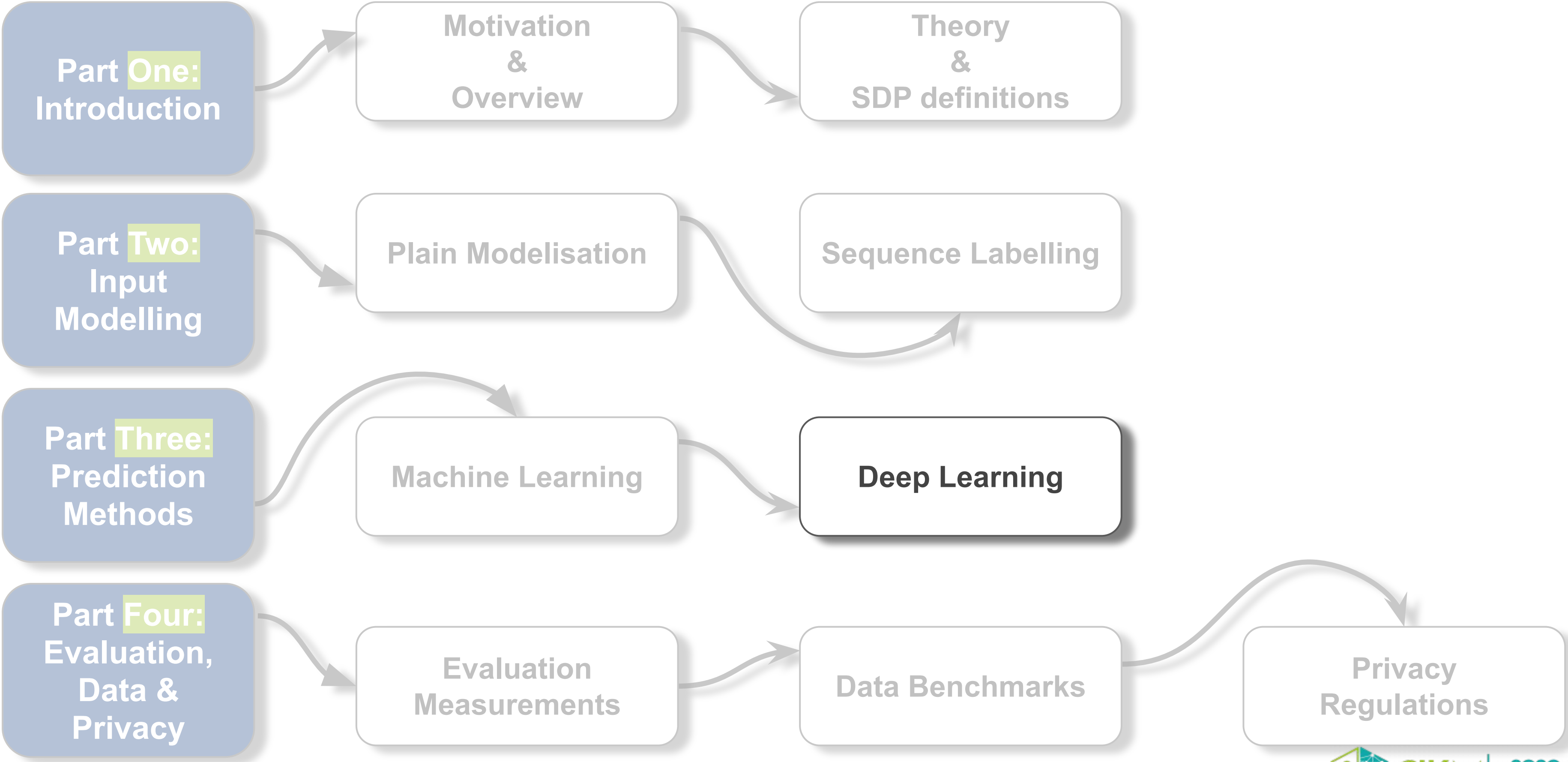


[14] Jing Chen, Jun Feng, Xia Sun, Nannan Wu, Zhengzheng Yang, and Sushing Chen. 2019. MOOC dropout prediction using a hybrid algorithm based on decision tree and extreme learning machine. Math. Probl. Eng. 2019 (2019).

Break Time

10 minutes

Roadmap



Part One: Introduction

Motivation & Overview

Theory & SDP definitions

Part Two: Input Modelling

Plain Modelisation

Sequence Labelling

Part Three: Prediction Methods

Machine Learning

Deep Learning

Part Four: Evaluation, Data & Privacy

Evaluation Measurements

Data Benchmarks

Privacy Regulations

Deep learning strategies

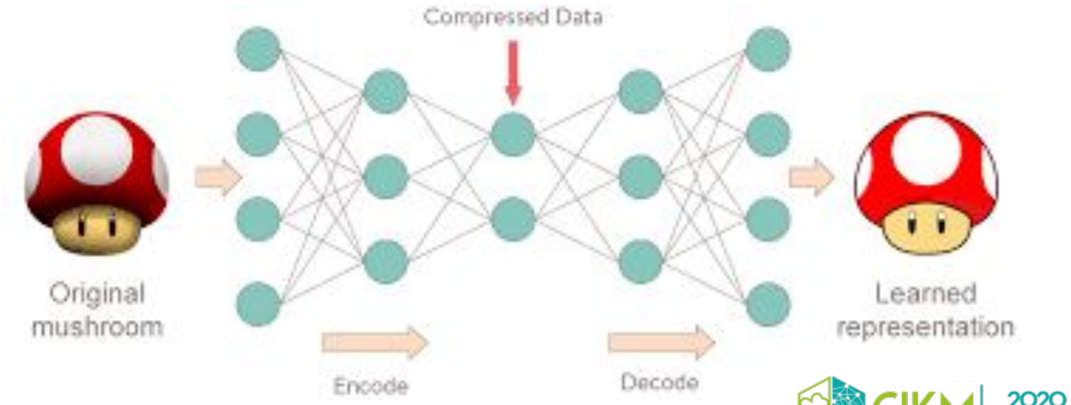
- Autoencoders (*preface*)
- Long-short Term Memory (LSTMs)
- Attention mechanism (*preface*)
- Convolutional neural networks (CNNs)
- RNNs + CNNs

Autoencoders

- Unsupervised technique for **representation learning**

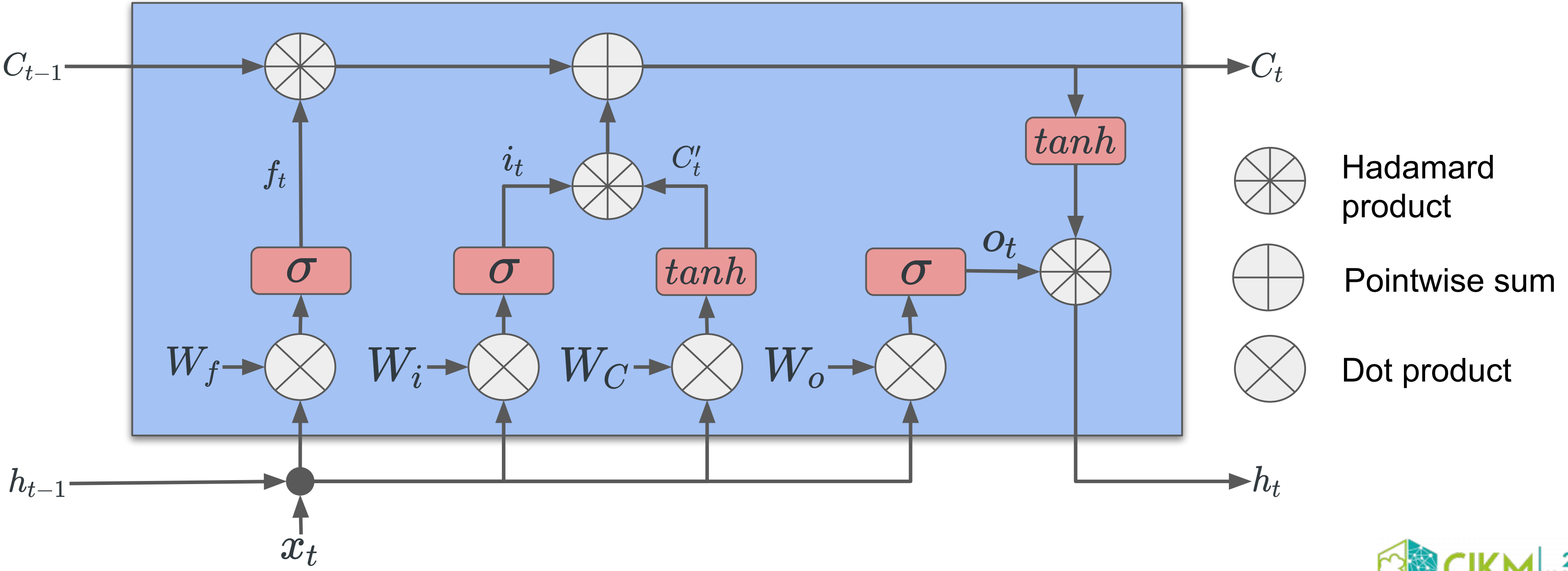
- Bottleneck in the network to compress the original input
  - If features are independent, *the task is difficult*
  - If there is ***structure***, learn it

- Applications
  - Anomaly detection
  - Data denoising
  - Image inpainting
  - Information retrieval



LSTMs

- E-tivities closer in time have a more dominant influence than those that are distant



LSTMs + AE (e.g. [22])

- Learn compact and effective representations from raw data

- Unsupervised approach to extract hidden info from the input sequences

- LSTM + AE for sequence-to-sequence learning

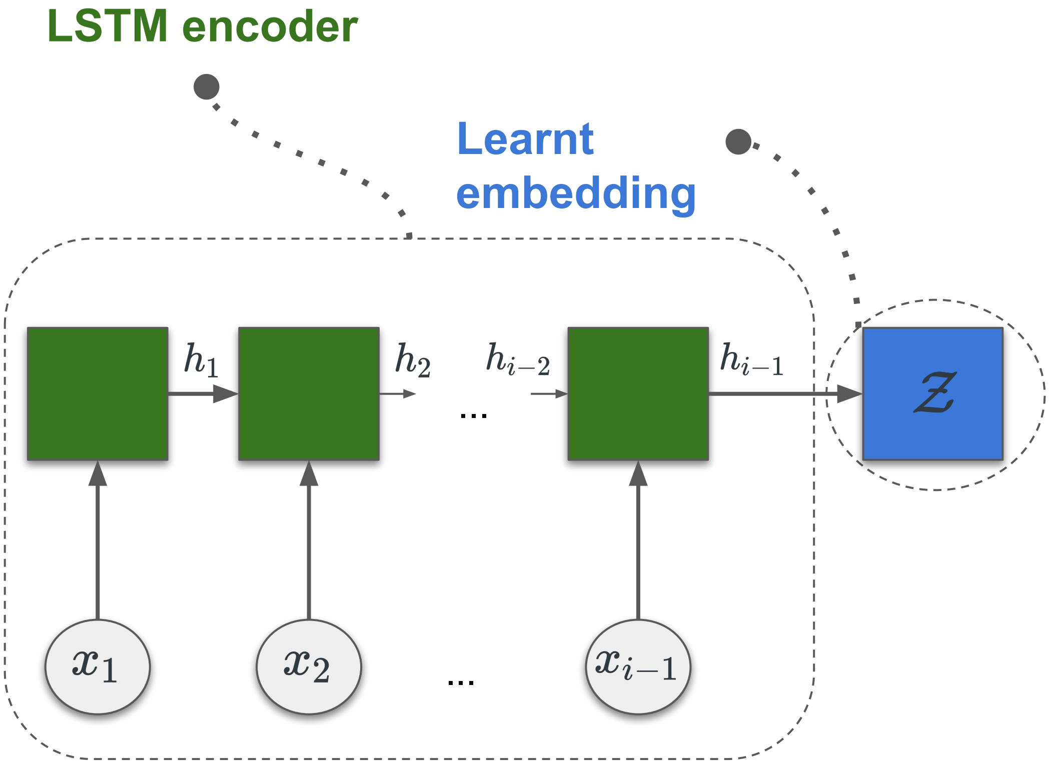
- **LSTM encoder + 2 LSTM decoders**

LSTMs + AE (e.g. [22])

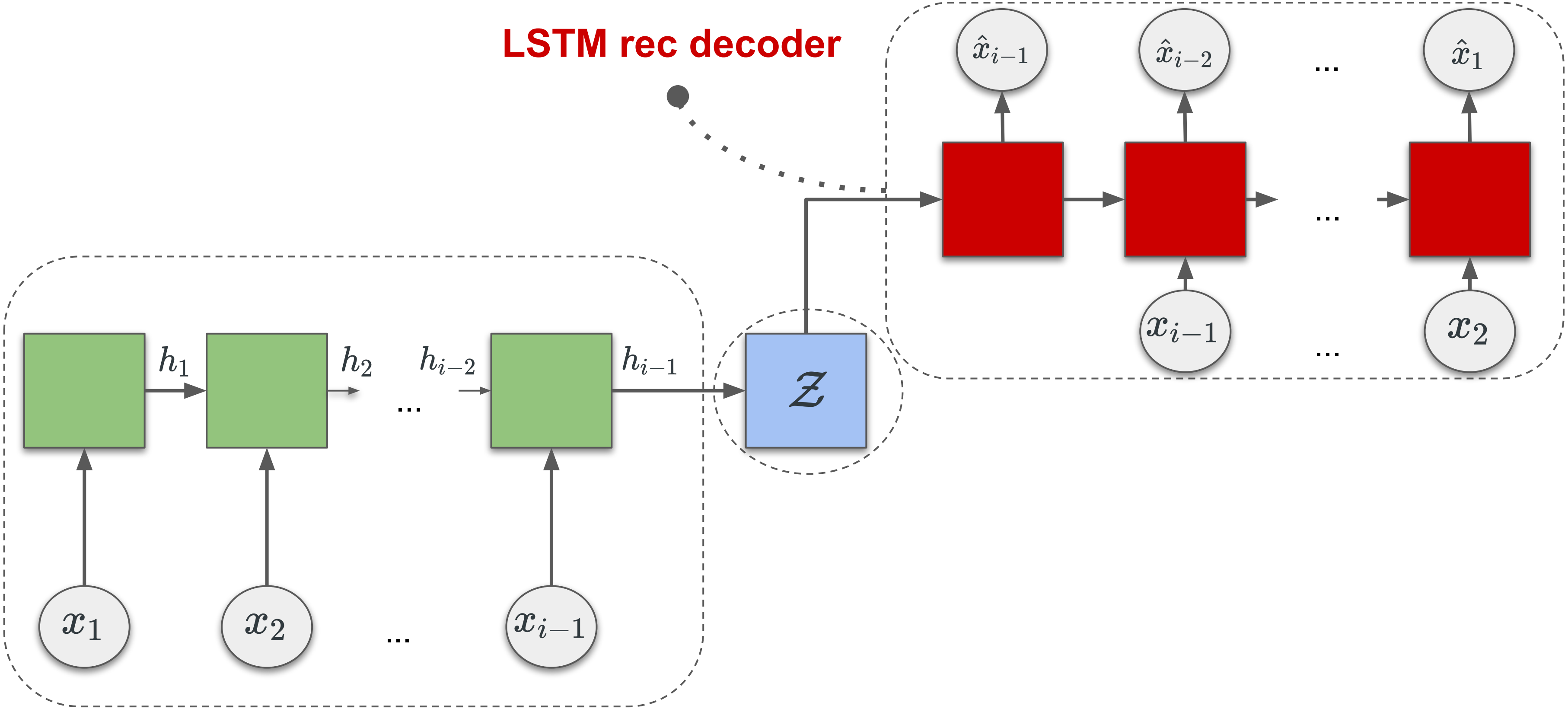
- Predict the student performance with behavioural patterns up to the i-th course phase

- Three components:
    - *LSTM encoder (encode up to i)*
    - *LSTM reconstruction decoder (decode up to i)*
    - *LSTM prediction decoder (decode up to k > i)*

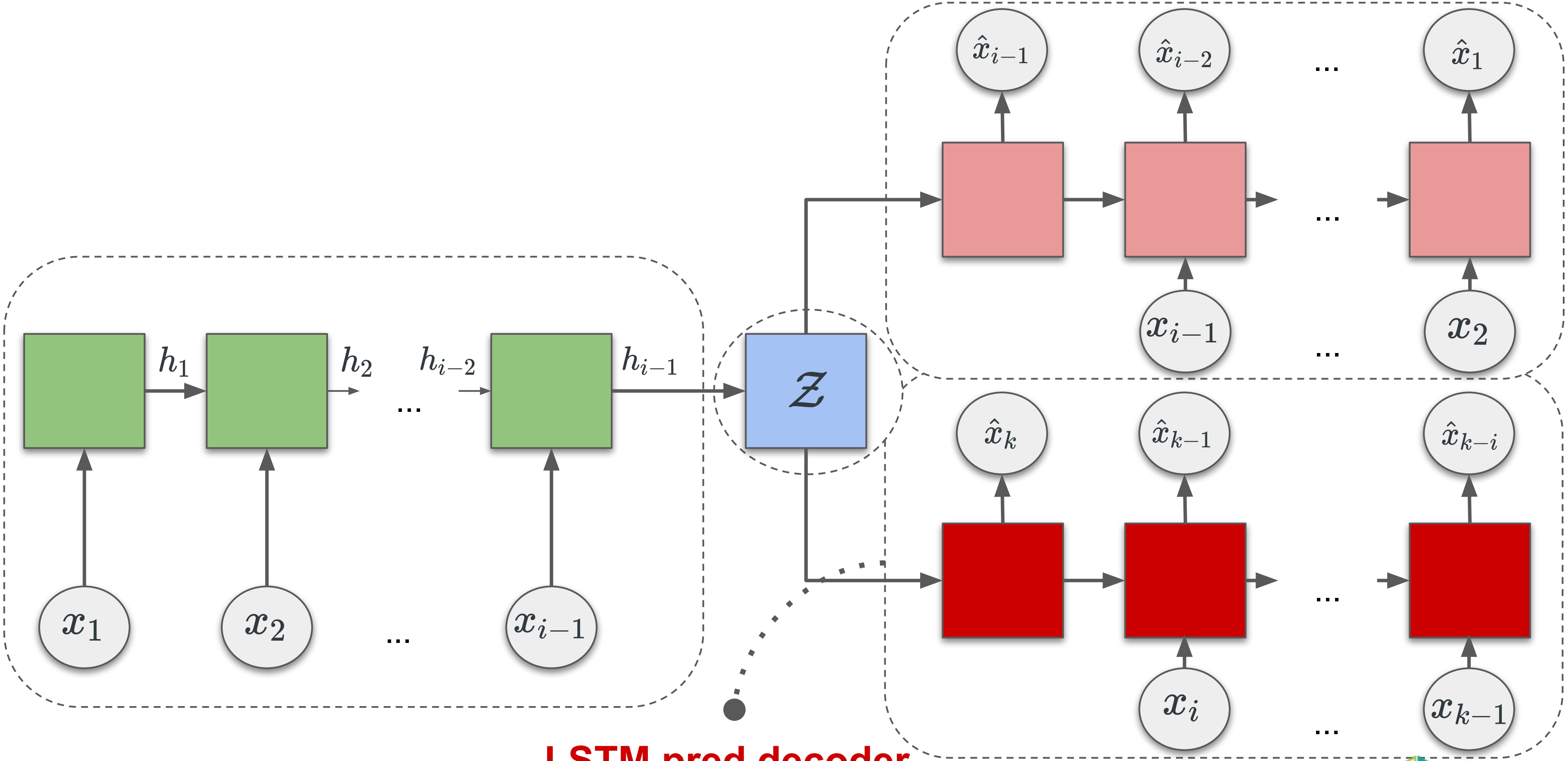
LSTMs + AE (e.g. [22])



LSTMs + AE (e.g. [22])

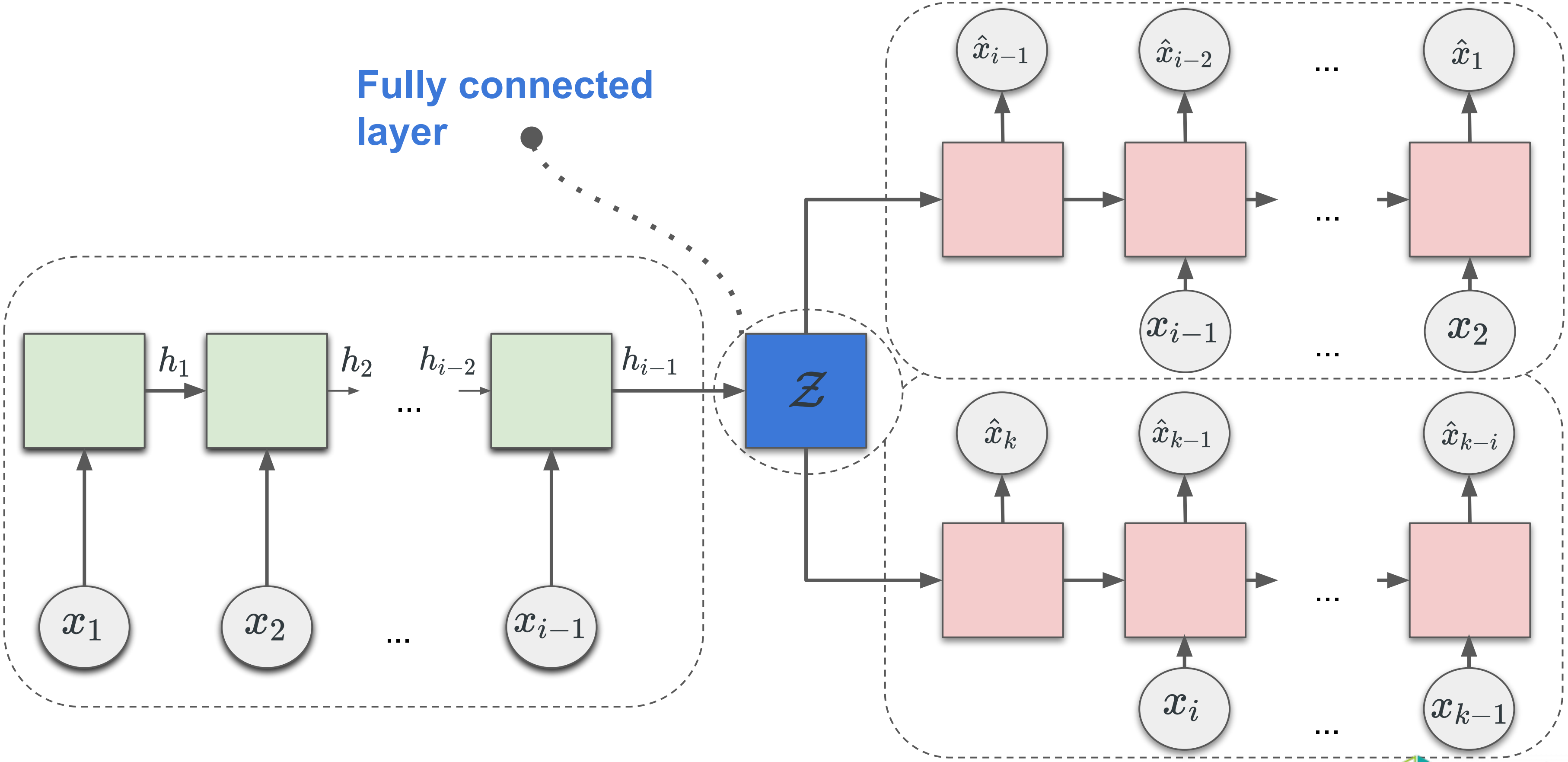


LSTMs + AE (e.g. [22])



LSTM pred decoder

LSTMs + AE (e.g. [22])



LSTMs + AE (e.g. [22])

- Loss function is Mean Square Error (**MSE**) between $\{x_1, \ldots, x_k\}$ and $\{\hat{x}_1, \ldots, \hat{x}_k\}$
- Encourage the model to use contextual locality as **1st priority**:
  - Assign **Gaussian weights** to each of the MSEs between true and reconstructed pairs

$$\mathcal{L} = \frac{1}{k} \sum_{n=1}^{k} e^{\frac{(i-n)^2}{2\sigma^2}} (x_n - \hat{x}_n)^2$$

Attention mechanism

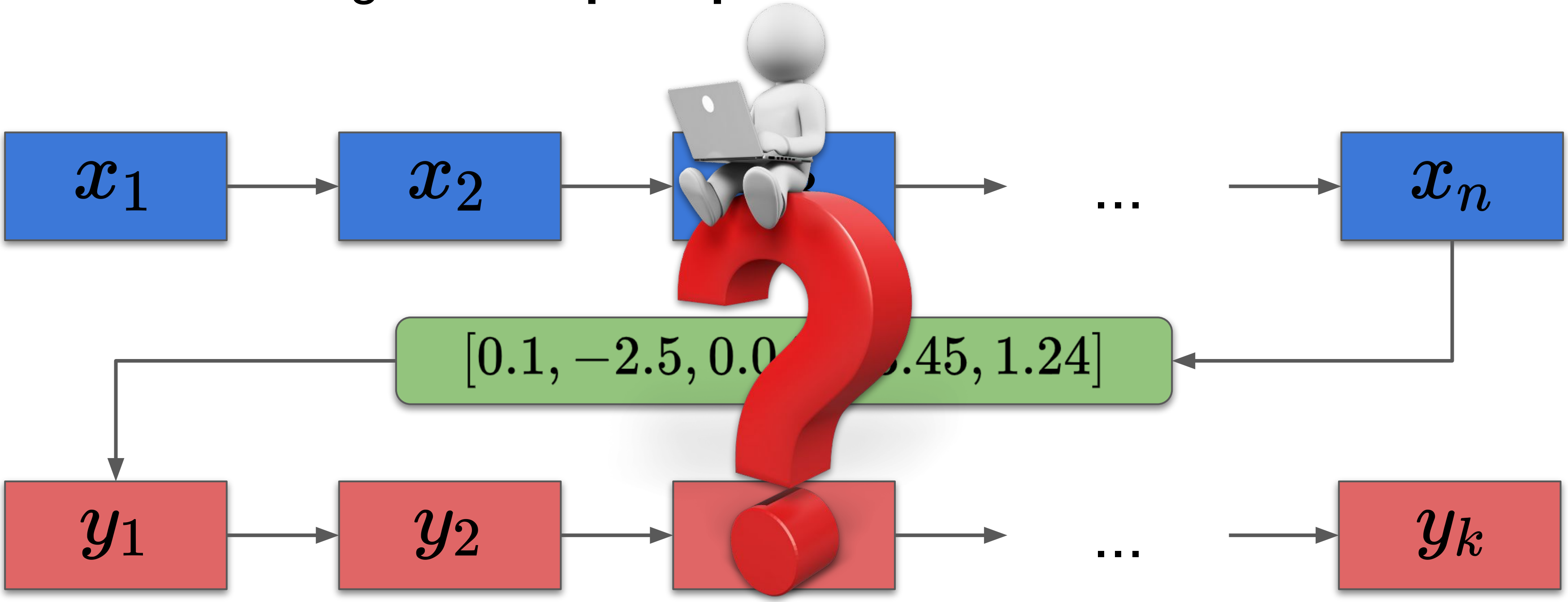
- Human visual attention focuses on certain regions
  - Look at priority first
  - Check less important features afterwards

- Humans can explain the relationship between words in a sentence

- **General idea**: interpret attention as a vector of importance weights according to the context

Attention mechanism [100]

- What's wrong with **Seq2Seq** model?

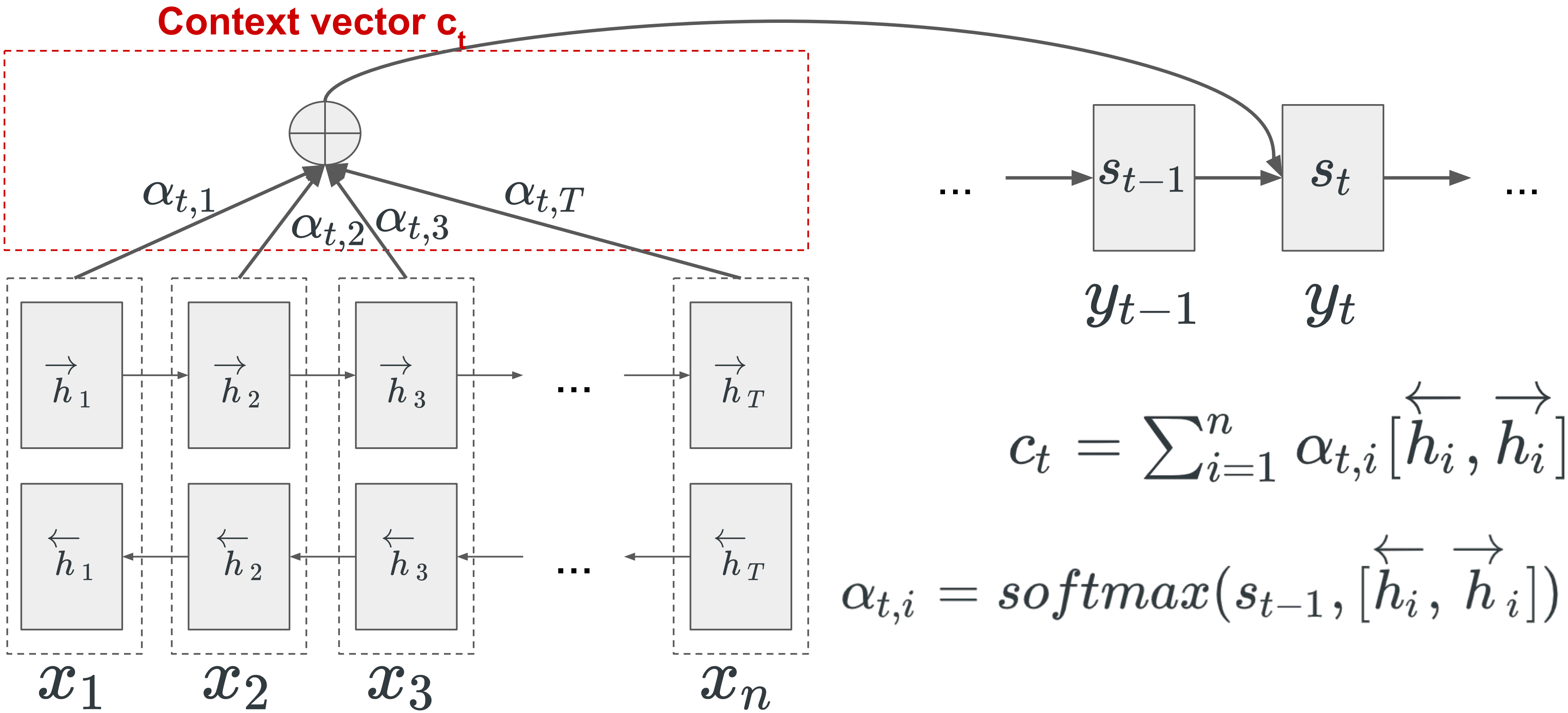


[100] Bahdanau, D., Cho, K. and Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Attention mechanism

- **Don’t build** a single context vector from the encoder’s last hidden state
- **Create shortcuts** between the context vector and the input sequence
- **Don’t worry about forgetting**:
  - The context vector **remembers** encoder/decoder hidden states, and the alignment between source and target

Attention mechanism



$$c_t = \sum_{i=1}^{n} \alpha_{t,i} [\overleftarrow{h_i}, \overrightarrow{h_i}]$$

$$\alpha_{t,i} = softmax(s_{t-1}, [\overleftarrow{h_i}, \overrightarrow{h}_i])$$

CNNs

- Use the convolutional operation to extract structural and temporal features of student behaviour

- **Combination with attention mechanism**: learn a context-aware representation for each e-tivity

- **Choice of using pooling layers or not**

CNNs [25]

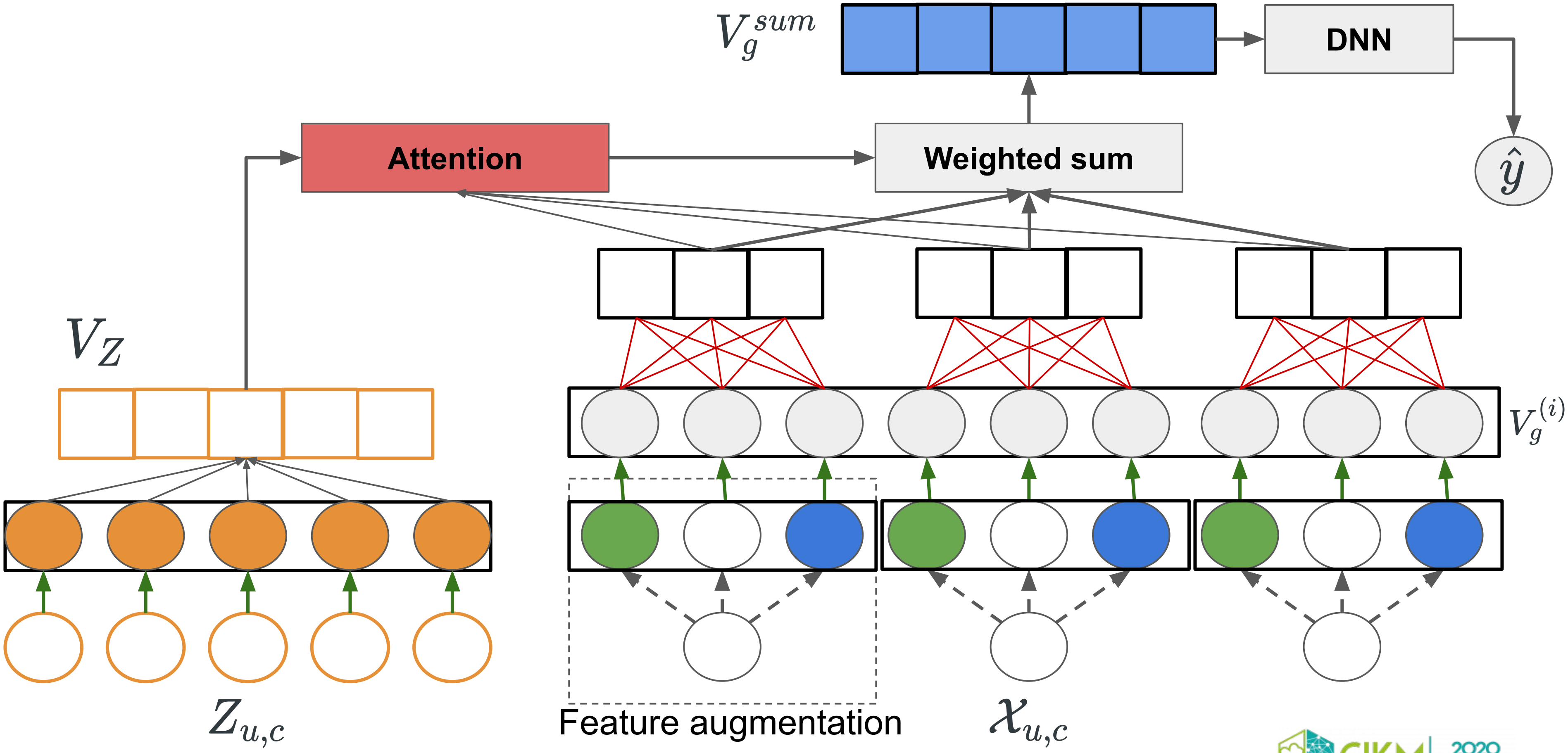
- **1-d convolution**

- Include context data
  - User and course information

- Three step training:
  - Context-smoothing - feature augmentation and embedding
  - Attention mechanism - **global attention**
  - Prediction component - dense neural network on the weighted-sum vector of the attention layer

[25] Wenzheng Feng, Jie Tang, and Tracy Xiao Liu. 2019. Understanding Dropouts in MOOCs. In AAAI 2019.

CNNs [25]


V_g^{sum}
DNN
$\hat{y}$
Attention
Weighted sum
V_Z
$V_g^{(i)}$
$Z_{u,c}$
Feature augmentation
$\mathcal{X}_{u,c}$

RNNs + CNNs [73]

- Get the best of both worlds
  - Spatial and temporal feature extraction

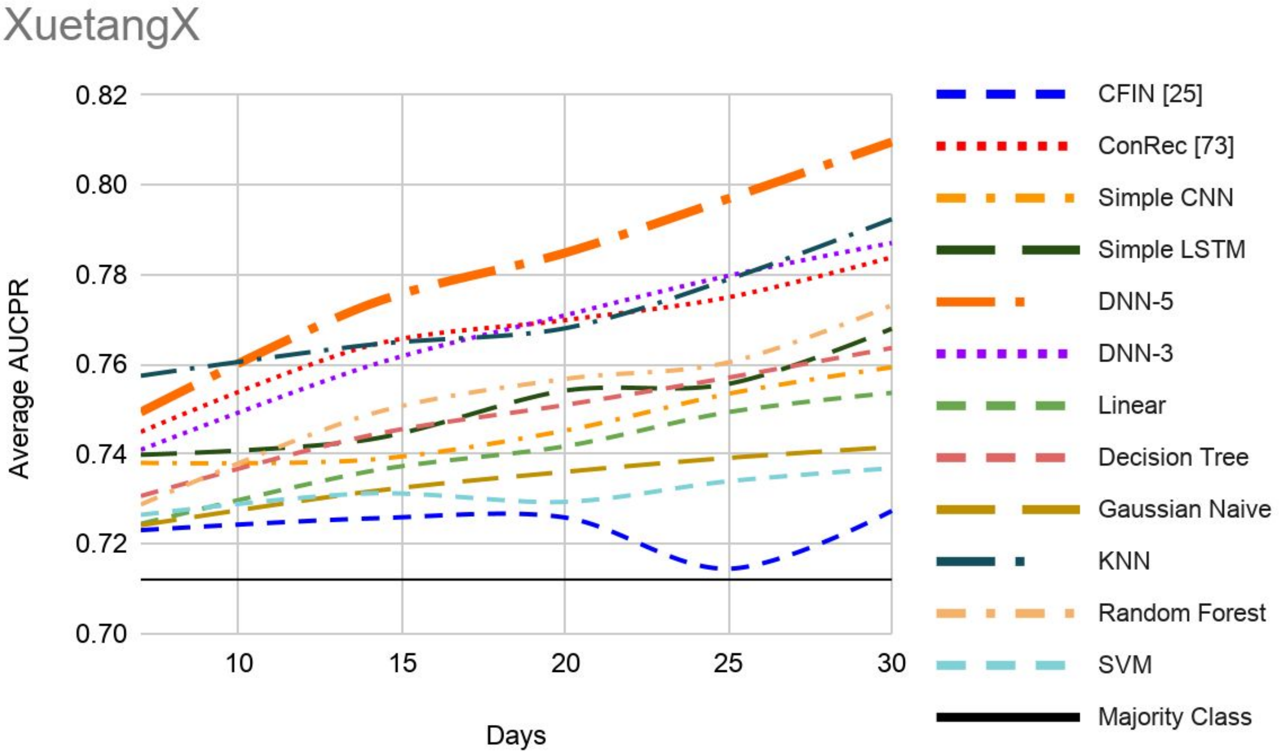
- CNN extracts features for the e-tivity matrices in each time slice

- Fuse the extracted features from the convolutional and pooling layers

- Give the fusion to the RNN to predict dropout/persister students

[25] Wei Wang, Han Yu, and Chuyan Miao. 2017. Deep model for dropout prediction in moocs. In Proceedings of the 2nd International Conference on Crowd Science and Engineering. ACM, ACM, New York, NY, USA, 26–32

Best performing method[1]

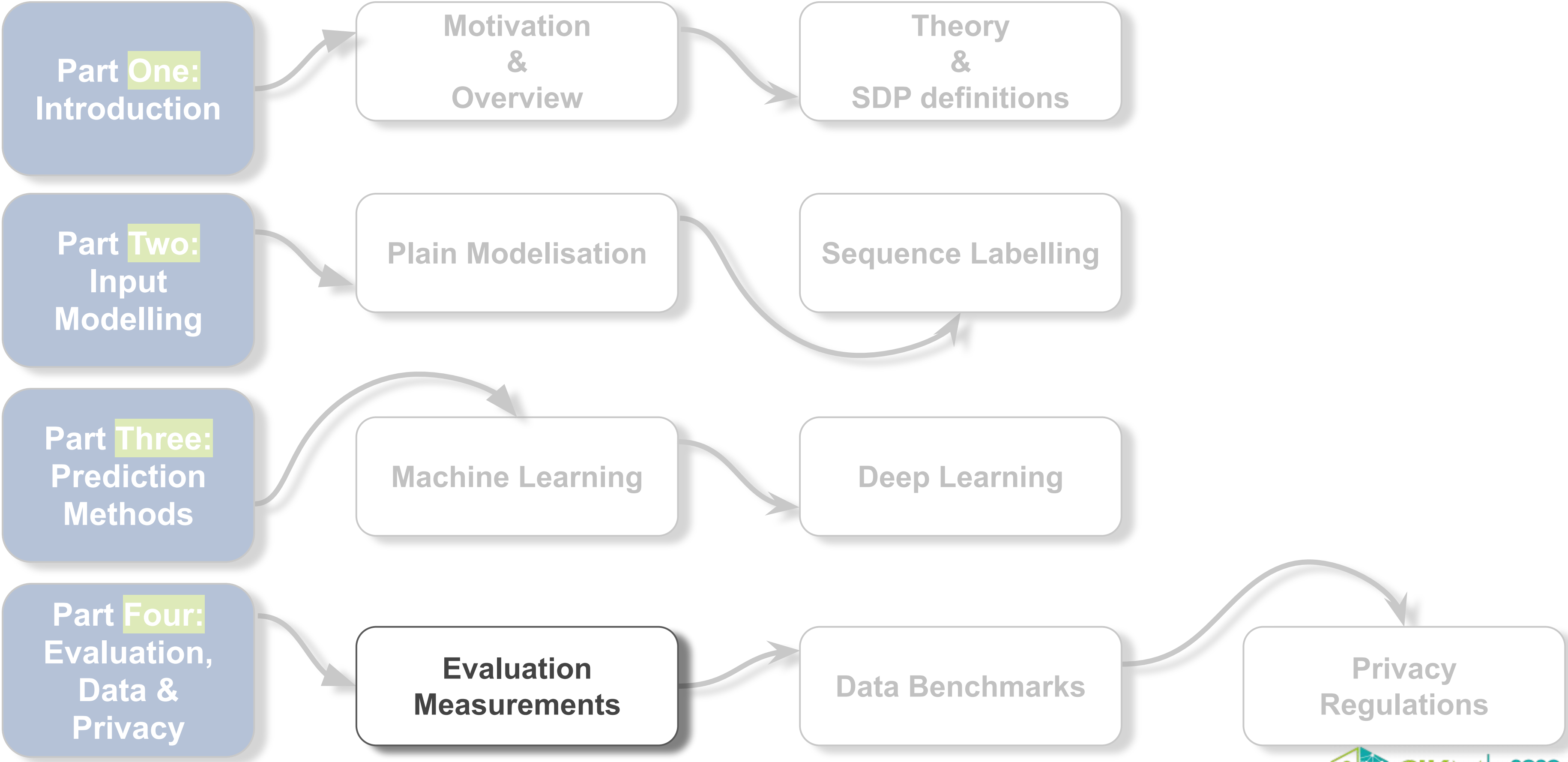


A reproducibility study of deep and surface machine learning methods for human-related trajectory prediction Authors: Bardh Prenkaj, Paola Velardi, Damiano Distante and Stefano Faralli. To be presented on the 21st of October at 8:05 AM - Waterford Zoom Room

Part Four

Evaluation, Data and Privacy

Roadmap



Part One: Introduction

Motivation & Overview

Theory & SDP definitions

Part Two: Input Modelling

Plain Modelisation

Sequence Labelling

Part Three: Prediction Methods

Machine Learning

Deep Learning

Part Four: Evaluation, Data & Privacy

Evaluation Measurements

Data Benchmarks

Privacy Regulations

Evaluation, datasets & privacy, open challenges

- Suitable evaluation measurements
- Lack of standard benchmarks
- Privacy regulations
- Data online publications
- Open challenges and future research in SDP

Evaluation measurements

- Dropout instances ≠ persister instances

- Data unbalancement

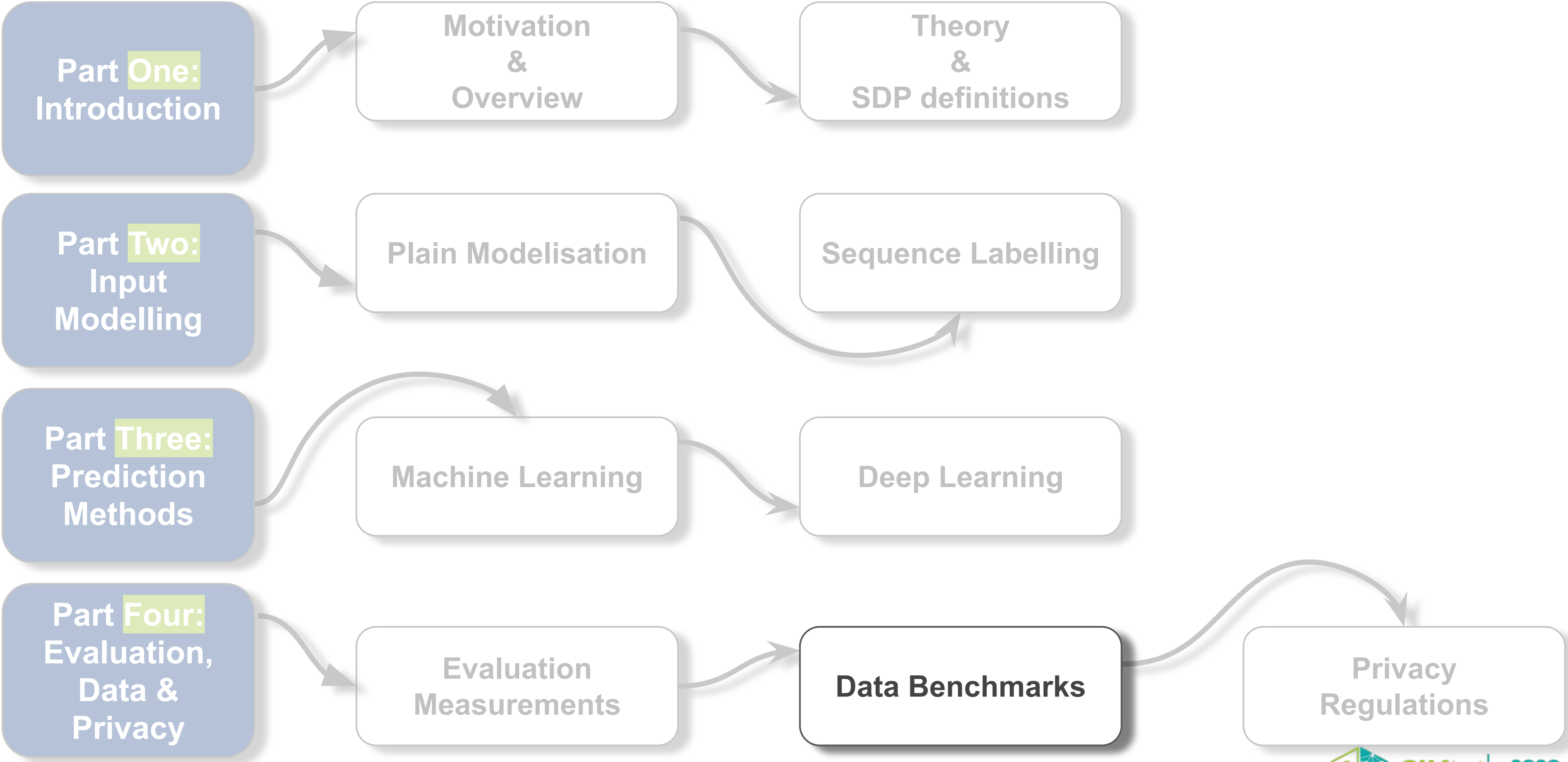
- Use evaluation metrics suitable to these problematics

- **Pioneer works used simple metrics (e.g. accuracy)**

Evaluation measurements

- Accuracy
- Recall, Precision, F1 score
- ROC plots
- PR plots
- AUC scores

Roadmap



Part One: Introduction

Motivation & Overview

Theory & SDP definitions

Part Two: Input Modelling

Plain Modelisation

Sequence Labelling

Part Three: Prediction Methods

Machine Learning

Deep Learning

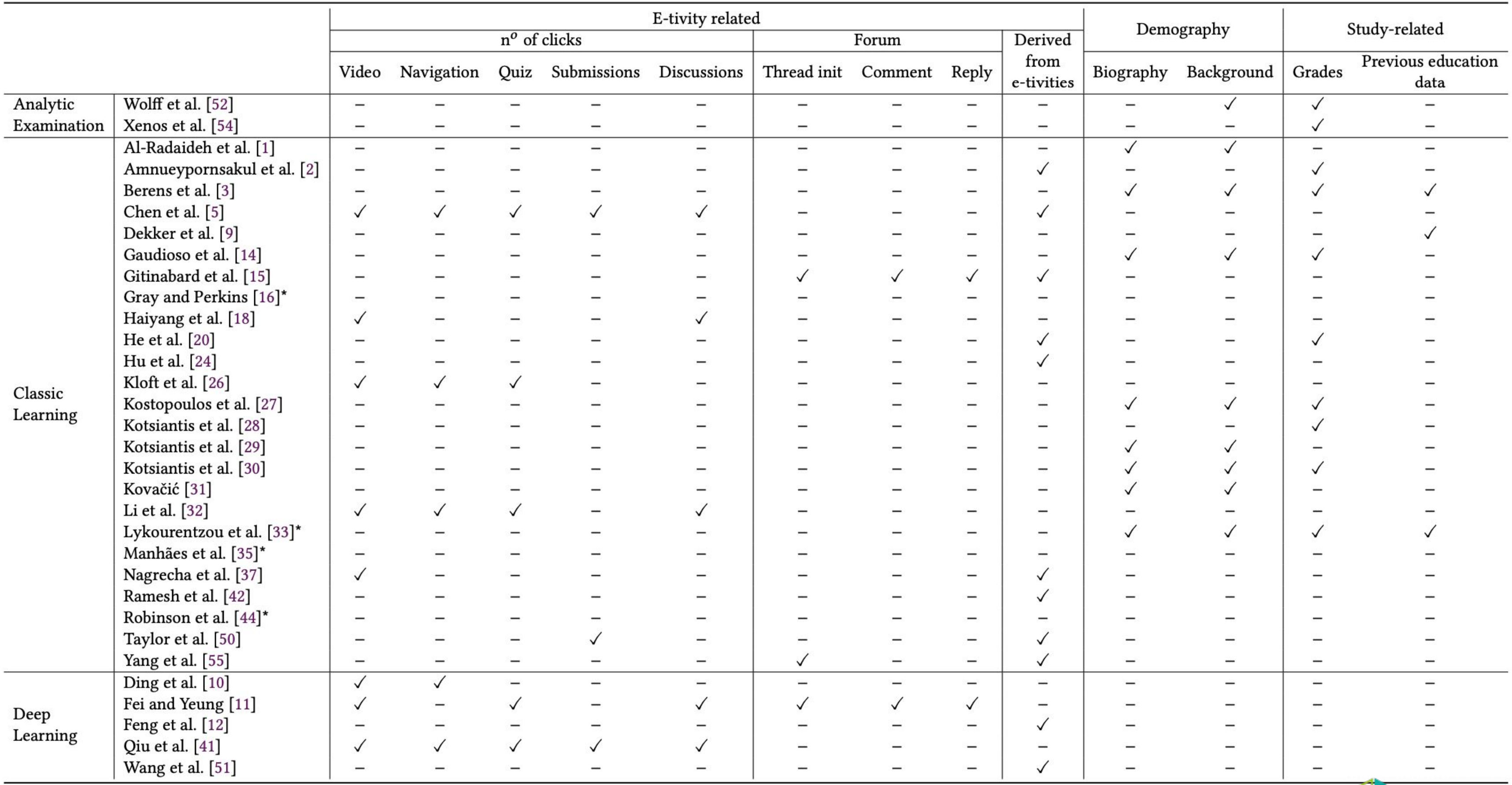
Part Four: Evaluation, Data & Privacy

Evaluation Measurements

Data Benchmarks

Privacy Regulations

Used features in the literature

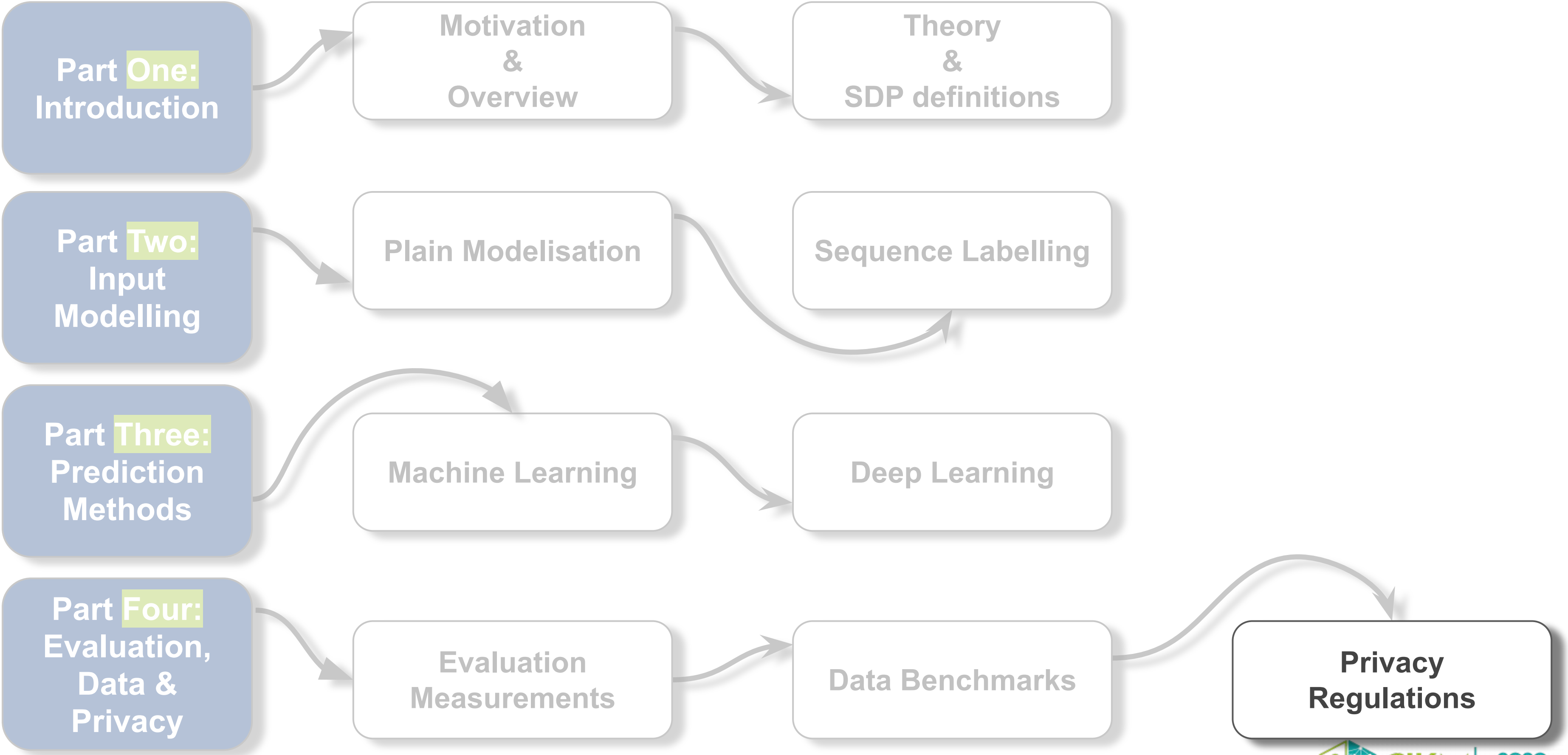


		E-tivity related									Demography		Study-related	
		n^o of clicks					Forum			Derived from e-tivities				
		Video	Navigation	Quiz	Submissions	Discussions	Thread init	Comment	Reply		Biography	Background	Grades	Previous education data
Analytic Examination	Wolff et al. [52]	–	–	–	–	–	–	–	–	–	–	✓	✓	–
	Xenos et al. [54]	–	–	–	–	–	–	–	–	–	–	–	✓	–
Classic Learning	Al-Radaideh et al. [1]	–	–	–	–	–	–	–	–	–	✓	✓	–	–
	Amnueypornsakul et al. [2]	–	–	–	–	–	–	–	–	✓	–	–	✓	–
	Berens et al. [3]	–	–	–	–	–	–	–	–	–	✓	✓	✓	✓
	Chen et al. [5]	✓	✓	✓	✓	✓	–	–	–	✓	–	–	–	–
	Dekker et al. [9]	–	–	–	–	–	–	–	–	–	–	–	–	✓
	Gaudioso et al. [14]	–	–	–	–	–	–	–	–	–	✓	✓	✓	–
	Gitinabard et al. [15]	–	–	–	–	–	✓	✓	✓	✓	–	–	–	–
	Gray and Perkins [16]*	–	–	–	–	–	–	–	–	–	–	–	–	–
	Haiyang et al. [18]	✓	–	–	–	✓	–	–	–	–	–	–	–	–
	He et al. [20]	–	–	–	–	–	–	–	–	✓	–	–	✓	–
	Hu et al. [24]	–	–	–	–	–	–	–	–	✓	–	–	–	–
	Kloft et al. [26]	✓	✓	✓	–	–	–	–	–	–	–	–	–	–
	Kostopoulos et al. [27]	–	–	–	–	–	–	–	–	–	✓	✓	✓	–
	Kotsiantis et al. [28]	–	–	–	–	–	–	–	–	–	–	–	✓	–
	Kotsiantis et al. [29]	–	–	–	–	–	–	–	–	–	✓	✓	–	–
	Kotsiantis et al. [30]	–	–	–	–	–	–	–	–	–	✓	✓	✓	–
	Kovačić [31]	–	–	–	–	–	–	–	–	–	✓	✓	–	–
	Li et al. [32]	✓	✓	✓	–	✓	–	–	–	–	–	–	–	–
	Lykourentzou et al. [33]*	–	–	–	–	–	–	–	–	–	✓	✓	✓	✓
	Manhães et al. [35]*	–	–	–	–	–	–	–	–	–	–	–	–	–
	Nagrecha et al. [37]	✓	–	–	–	–	–	–	–	✓	–	–	–	–
	Ramesh et al. [42]	–	–	–	–	–	–	–	–	✓	–	–	–	–
	Robinson et al. [44]*	–	–	–	–	–	–	–	–	–	–	–	–	–
	Taylor et al. [50]	–	–	–	✓	–	–	–	–	✓	–	–	–	–
	Yang et al. [55]	–	–	–	–	–	✓	–	–	✓	–	–	–	–
Deep Learning	Ding et al. [10]	✓	✓	–	–	–	–	–	–	–	–	–	–	–
	Fei and Yeung [11]	✓	–	✓	–	✓	✓	✓	✓	–	–	–	–	–
	Feng et al. [12]	–	–	–	–	–	–	–	–	✓	–	–	–	–
	Qiu et al. [41]	✓	✓	✓	✓	✓	–	–	–	–	–	–	–	–
	Wang et al. [51]	–	–	–	–	–	–	–	–	✓	–	–	–	–

Standard Benchmarks

- Time-series and time-related information
  - XuetangX
  - KDDCup15
- Time-agnostic and student background information
  - HarvardX

Roadmap



Part One: Introduction

Motivation & Overview

Theory & SDP definitions

Part Two: Input Modelling

Plain Modelisation

Sequence Labelling

Part Three: Prediction Methods

Machine Learning

Deep Learning

Part Four: Evaluation, Data & Privacy

Evaluation Measurements

Data Benchmarks

Privacy Regulations

Privacy Regulations

- **Data** coming from **institutions** or well-established and **notorious MOOCs** (e.g. Coursera, edX, and Udemy). Typically the legal **responsibilities** to the body of the data **property**;

- Data **privacy** is becoming a **primary** concern **worldwide**;

- The **US lacks** a comprehensive **federal law** that regulates the collection and use of personal information. Instead, the **government** has **approached** privacy and security by **adjusting** only **actors** and **types** of sensitive **information**, creating overlapping and **contradictory protections**;

- The **EU** General Data Protection Regulation (**GDPR**) intends to:
  - **Harmonise** data privacy **laws** across Europe.
  - **Protect** and **empower** all EU citizens data privacy.
  - **Reshape** the way organisations across the region **approach** data privacy

Privacy-compliant dataset publication

- ***Students*** with their **personal** information generally **anonymised**.
  Notice that the **interaction** among them should be **protected** as well.

- ***Courses*** are allotted in course phases, each of whom **contains** several **resources** (e.g. videos, quizzes, reading lectures).
- ***E-tivities*** are the activities produced by events that **students** cause during the **interaction** with the e-platform.

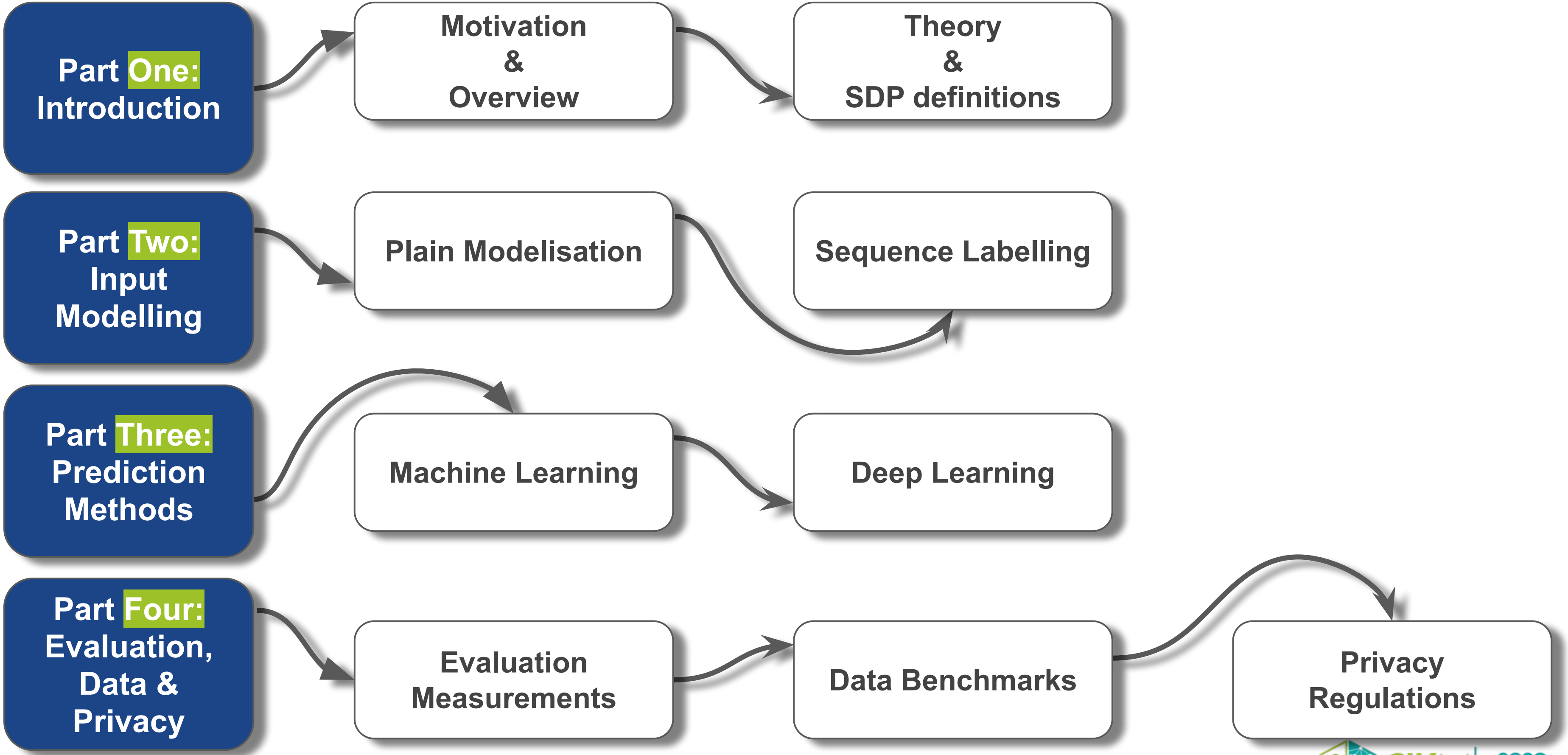
- **Privacy-Preserving Data Mining (PPDM)**, a multi-disciplinary research area, consists of:
  - transforms the original data so that the data mining processes do not violate privacy constraints:
    - Randomisation;
    - Anonymisation;
    - Encryption;
- For a detailed review of the previous strategies, we point to [81]

[81] Aobakwe Senosi and George Sibiya. 2017. Classification and evaluation of privacy preserving data mining: a review. In 2017 IEEE AFRICON. IEEE, 849–855.

Part Five

Conclusion

Roadmap



Part One: Introduction

Motivation & Overview

Theory & SDP definitions

Part Two: Input Modelling

Plain Modelisation

Sequence Labelling

Part Three: Prediction Methods

Machine Learning

Deep Learning

Part Four: Evaluation, Data & Privacy

Evaluation Measurements

Data Benchmarks

Privacy Regulations

Open Challenges

- Adopt better common benchmarks:
    - Exist difficulties in comparing different approaches;
        - **absence** of standard **benchmarks**;
        - Identifies an index of available **datasets**;
        - set **common** evaluation **metrics** and baselines;
        - systematic organisation of public **challenges**.

- Deep sequential methods should be better explored:
    - **information** from previous course **phases** is **relevant**
    - **Deep** learning methods are better suited for **sequence labelling**;
    - Deep learning in SDP is still **not** sufficiently **explored**;
    - Deep methods, are **poor interpretability**
      (Interpretability is critical in the e-learning field, to prescribe effective prevention strategies);

Open Challenges

- Online degree peculiarities:
    - literature does **not** focus on **online degree** (fast-paced and short-term MOOCs);
    - universities offer online degree courses, **sophisticated models** are **needed**;
    - abandon may arise from **complex interactions** (sequential and **parallel** activities);

- Duration of e-tivity problems:
    - literature does **not** consider **temporal lags** on completing e-tivities;
    - **modelling intervals** the amount of time that a student **engages**;
    - **crucial** for corporate **universities**;
    - **inter-stage** gaps that can be **tailored ad-hoc** for each **student** (personalization).

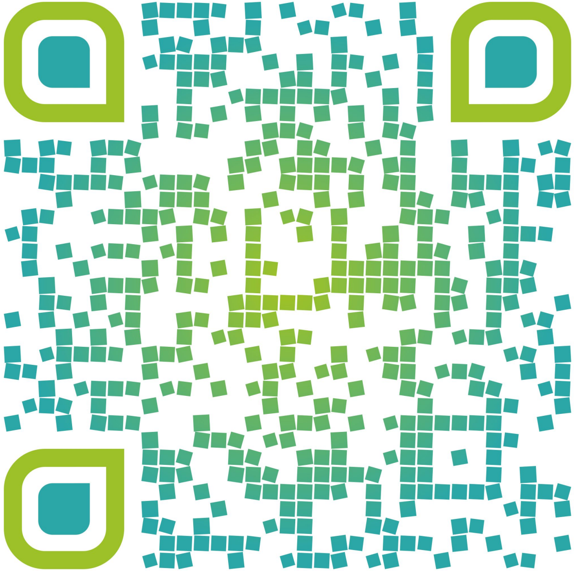
thank you!

Bardh Prenkaj

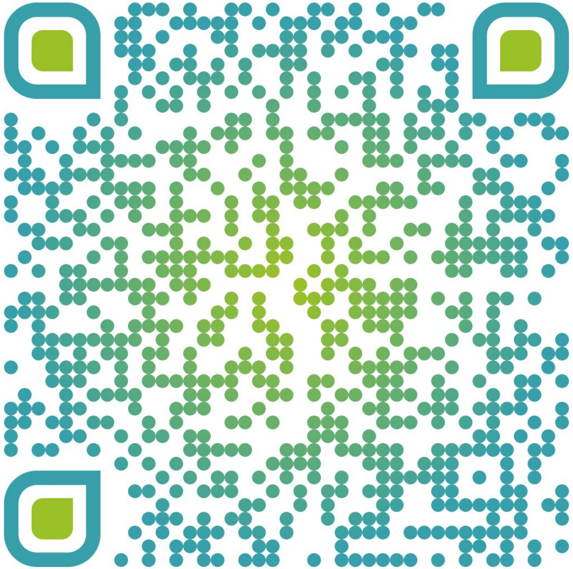
Giovanni Stilo

Lorenzo Madeddu

Online Resources



Tutorials Website



Tutorials Slides

Bibliography (1)

[1] Qasem A. Al-Radaideh, Emad M. Al-Shawakfa, and Mustafa I. Al-Najjar. 2006. Mining student data using decision trees. In International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan. 1–5.

[2] Sattar Ameri, Mahtab J. Fard, Ratna B. Chinnam, and Chandan K. Reddy. 2016. Survival analysis based framework for early prediction of student dropouts. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 903–912.

[3] Bussaba Amnueypornsakul, Suma Bhat, and Phakpoom Chinprutthiwong. 2014. Predicting attrition along the way: The UIUC model. In Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs. Association for Computational Linguistics, Doha, Qatar, 55–59.

[4] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014).

[5] Behdad Bakhshinategh, Osmar R Zaiane, Samira ElAtia, and Donald Ipperciel. 2018. Educational datamining applications and tasks: A survey of the last 10 years. Education and Information Technologies 23, 1 (2018), 537–553.

[6] V.K. Balakrishnan. 1997. Schaum's Outline of Graph Theory: Including Hundreds of Solved Problems. McGraw Hill Professional, New York, NY, USA.

[7] Papia Bawa. 2016. Retention in online courses: Exploring issues and solutions—A literature review. Sage Open 6, 1 (2016), 2158244015621777.

[8] Johannes Berens, Kerstin Schneider, Simon Görtz, Simon Oster, and Julian Burghoff. 2018. Early Detection of Students at Risk–Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods. CESifo Working Paper, Munich, Germany.

[9] Leo Breiman. 2017. Classification and regression trees. Routledge, Abingdon, Oxfordshire, UK.

[10] Peter J. Brockwell, Richard A. Davis, and Matthew V. Calder. 2002. Introduction to time series and forecasting. Vol. 2. Springer, New York, NY, USA.

[11] Rebecca Brown, Collin Lynch, Yuan Wang, Michael Eagle, Jennifer Albert, Tiffany Barnes, Ryan Shaun Baker, Yoav Bergner, and Danielle S. McNamara. 2015. Communities of Performance & Communities of Preference. In CEUR Workshop Proceedings, Vol. 1446. CEUR-WS, USA.

[12] Vicki Carter. 1996. Do media influence learning? Revisiting the debate in the context of distance education. Open Learning: The Journal of Open, Distance and e-Learning 11, 1 (1996), 31–40.

[13] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. 2010. Semi-Supervised Learning. The MIT Press, Cambridge, MA, USA.

[14] Jing Chen, Jun Feng, Xia Sun, Nannan Wu, Zhengzheng Yang, and Sushing Chen. 2019. MOOC Dropout Prediction Using a Hybrid Algorithm Based on Decision Tree and Extreme Learning Machine. Mathematical Problems in Engineering 2019 (2019).

[15] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, ACM, New York, NY, USA, 785–794.

[16] Yujing Chen, Aditya Johri, and Huzefa Rangwala. 2018. Running out of stem: a comparative study across stem majors of college students at-risk of dropping out early. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge. ACM, 270–279.

[17] David R. Cox. 1972. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) 34, 2 (1972), 187–202.

[18] Fisnik Dalipi, Ali Shariq Imran, and Zenun Kastrati. 2018. MOOC dropout prediction using machine learning techniques: Review and research challenges. In 2018 IEEE Global Engineering Education Conference (EDUCON). IEEE, 1007–1014.

[19] Jesse Davis and Mark Goadrich. 2006. The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd international conference on Machine learning. ACM, ACM, New York, NY, USA, 233–240.

Bibliography (2)

[20] Gerben W. Dekker, Mykola Pechenizkiy, and Jan M. Vleeshouwer. 2009. Predicting Students Drop Out: A Case Study. International Working Group on Educational Data Mining (2009).

[21] David P. Diaz. 2000. Comparison of student characteristics, and evaluation of student success, in an online health education course. Ph.D. Dissertation. Nova Southeastern University.

[22] Mucong Ding, Kai Yang, Dit-Yan Yeung, and Ting-Chuen Pong. 2018. Effective Feature Learning with Unsupervised Learning for Improving the Predictive Models in Massive Open Online Courses. arXiv:1812.05044

[23] William Doherty. 2006. An analysis of multiple factors affecting retention in web-based community college courses. The Internet and Higher Education 9, 4 (2006), 245–255.

[24] Mi Fei and Dit-Yan Yeung. 2015. Temporal models for predicting student dropout in massive open online courses. In 2015 IEEE International Conference on Data Mining Workshop (ICDMW). IEEE, 256–263.

[25] Wenzheng Feng, Jie Tang, and Tracy Xiao Liu. 2019. Understanding Dropouts in MOOCs. In AAAI 2019.

[26] Karen Frankola. 2001. Why online learners dropout. WORKFORCE-COSTA MESA-80, 10 (2001), 52–61. http://www.workforce.com/feature/00/07/29

[27] S Hari Ganesh and A Joy Christy. 2015. Applications of educational data mining: a survey. In 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). IEEE, 1–6.

[28] Josh Gardner and Christopher Brooks. 2018. Student success prediction in MOOCs. User Modeling and User-Adapted Interaction 28,2(2018),127–203.

[29] Elena Gaudioso, Miguel Montero, and Felix Hernandez-Del-Olmo. 2012. Supporting teachers in adaptive educational systems through predictive models: A proof of concept. Expert Systems with Applications 39, 1 (2012), 621–625.

[30] Niki Gitinabard, Farzaneh Khoshnevisan, Collin F. Lynch, and Elle Yuan Wang. 2018. Your Actions or Your Associates? Predicting Certification and Dropout in MOOCs with Behavioral and Social Features. arXiv:1809.00052

[31] Cameron C. Gray and Dave Perkins. 2019. Utilizing early engagement and machine learning to predict student outcomes. Computers & Education 131 (2019), 22–32.

[32] Liu Haiyang, Zhihai Wang, Phillip Benachour, and Philip Tubman. 2018. A Time Series Classification Method for Behaviour-Based Dropout Prediction. In 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT). IEEE, 191–195.

[33] Jiazhen He,James Bailey, BenjaminI. P. Rubinstein, and Rui Zhang. 2015. Identifying at-risk students in massive open online courses. In Twenty-Ninth AAAI Conference on Artificial Intelligence.

[34] Michael Herbert. 2006. Staying the course: A study in online student satisfaction and retention. Online Journal of Distance Learning Administration 9, 4 (2006), 300–317.

[35] Erin Heyman. 2010. Overcoming student retention issues in higher education online programs. Online Journal of Distance Learning Administration 13, 4 (2010).

[36] Deng Houtao, Runger C. George, Tuv Eugene, and Martyanov Vladimir. 2013. A Time Series Forest for Classification and Feature Extraction. Inf. Sci. 239 (2013), 142–153.

[37] Ya-Han Hu, Chia-Lun Lo,and Sheng-Pao Shih. 2014. Developing early warning systems to predict students'online learning performance. Computers in Human Behavior 36 (2014), 469–478.

Bibliography (3)

[38] Gordon V. Kass. 1980. An exploratory technique for investigating large quantities of categorical data. Journal of the Royal Statistical Society: Series C (Applied Statistics) 29, 2 (1980), 119–127.

[39] Tom Kasuba. 1993. Simplified fuzzy ARTMAP. AI Expert (1993).

[40] Usha Keshavamurthy and H. S. Guruprasad. 2014. Learning Analytics: A Survey. International Journal of Computer Trends and Technology (IJCTT) 18(6) (2014).

[41] Marius Kloft, Felix Stiehler, Zhilin Zheng, and Niels Pinkwart. 2014. Predicting MOOC dropout over weeks using machine learning methods. In Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs. 60–65.

[42] Georgios Kostopoulos, Sotiris Kotsiantis, and Panagiotis Pintelas. 2015. Estimating student dropout in distance higher education using semi- supervised techniques. In Proceedings of the 19th Panhellenic Conference on Informatics. ACM, ACM, New York, NY, USA, 38–43.

[43] Sotiris Kotsiantis, Kiriakos Patriarcheas, and Michalis Xenos. 2010. A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. Knowledge-Based Systems 23, 6 (2010), 529–535.

[44] Sotiris Kotsiantis, Christos Pierrakeas, and Panagiotis Pintelas. 2003. Preventing student dropout in distance learning using machine learning techniques. In International conference on knowledge-based and intelligent information and engineering systems. Springer, New York, NY, USA, 267–274.

[45] Sotiris Kotsiantis, Christos Pierrakeas, Ioannis Zaharakis, and Panagiotis Pintelas. 2003. Efficiency of machine learning techniques in predicting students performance in distance learning systems. University of Patras Press, 297–306.

[46] Zlatko J. Kovačić. 2010. Early prediction of student success: Mining student enrollment data. In Proceedings of Informing Science & IT Education Conference. Citeseer.

[47] George D. Kuh. 2009. The National Survey of Student Engagement: Conceptual and empirical foundations. New Directions for Institutional Research 2009(122009),5–20. https://doi.org/10.1002/ir.283

[48] Anupama S. Kumar and MN Vijayalakshmi. 2012. Mining of student academic evaluation records in higher education. In 2012 International Conference on Recent Advances in Computing and Software Systems. IEEE, 67–70.

[49] Mukesh Kumar, AJ Singh, and Disha Handa. 2017. Literature survey on educational dropout prediction. International Journal of Education and Management Engineering 7, 2 (2017), 8.

[50] Wentao Li, Min Gao, Hua Li, Qingyu Xiong, Junhao Wen, and Zhongfu Wu. 2016. Dropout prediction in MOOCs using behavior features and multi-view semi-supervised learning. In 2016 international joint conference on neural networks (IJCNN). IEEE, 3130–3137.

[51] Nick Littlestone and Manfred K. Warmuth. 1994. The weighted majority algorithm. Information and computation 108, 2 (1994), 212–261.

[52] Chu Kiong Loo and MVC Rao. 2005. Accurate and reliable diagnosis and classification using probabilistic ensemble simplified fuzzy ARTMAP. IEEE Transactions on Knowledge and Data engineering 17, 11 (2005), 1589–1593.

[53] Ioanna Lykourentzou, Ioannis Giannoukos, Vassilis Nikolopoulos, George Mpardis, and Vassili Loumos. 2009. Dropout prediction in e-learning courses through the combination of machine learning techniques. Computers & Education 53, 3 (2009), 950–965.

[54] Laci Mary Barbosa Manhães, Sérgio Manuel Serrada Cruz, and Geraldo Zimbrão. 2014. WAVE: an architecture for predicting dropout in undergraduate courses using EDM. In Proceedings of the 29th Annual ACM Symposium on Applied Computing. ACM, ACM, New York, NY, USA, 243–247.

[55] Mary McHugh. 2012. Interrater reliability: The kappa statistic. Biochemia medica : časopis Hrvatskoga društva medicinskih biokemičara / HDMB 22 (102012),276–282. https://doi.org/10.11613/BM.2012.031

Bibliography (4)

[56] Othon Michail. 2016. An introduction to temporal graphs: An algorithmic perspective. Internet Mathematics 12, 4 (2016), 239–280.

[57] Christoph Molnar. 2018. Interpretable Machine Learning. github. io/interpretable-ml-book. (2018).

[58] Michael Morgan, Matthew Butler, Neena Thota, and Jane Sinclair. 2018. How CS academics view student engagement. In Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education. ACM, ACM, New York, NY, USA, 284–289.

[59] Saurabh Nagrecha, John Z. Dillon, and Nitesh V. Chawla. 2017. MOOC dropout prediction: lessons learned from making pipelines interpretable. In Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, 351–359.

[60] Alejandro Peña-Ayala. 2014. Educational data mining: A survey and a data mining-based analysis of recent works. Expert systems with applications 41, 4 (2014), 1432–1462.

[61] Bardh Prenkaj, Paola Velardi, Giovanni Stilo, Damiano Distante, and Stefano Faralli. 2020. A Survey of Machine Learning Approaches for Student Dropout Prediction in Online Courses. 53, 3, Article 57 (May 2020), 34 pages. https://doi.org/10.1145/3388792

[62] Jiezhong Qiu, Jie Tang, Tracy Xiao Liu, Jie Gong, Chenhui Zhang, Qian Zhang, and Yufei Xue. 2016. Modeling and predicting learning behavior in MOOCs. In Proceedings of the ninth ACM international conference on web search and data mining. ACM, ACM, New York, NY, USA, 93–102.

[63] Lin Qiu, Yanshen Liu, Quan Hu, and Yi Liu. 2018. Student dropout prediction in massive open online courses by convolutional neural networks. Soft Computing (2018), 1–15.

[64] Ross J. Quinlan. 2014. C4.5: programs for machine learning. Elsevier, New York, NY, USA.

[65] Arti Ramesh, Dan Goldwasser, Bert Huang, Hal Daume III, and Lise Getoor. 2014. Learning latent engagement patterns of students in online courses. In Twenty-Eighth AAAI Conference on Artificial Intelligence.

[66] Carly Robinson, Michael Yeomans, Justin Reich, Chris Hulleman, and Hunter Gehlbach. 2016. Forecasting student achievement in MOOCs with natural language processing. In Proceedings of the sixth international conference on learning analytics & knowledge. ACM, ACM, New York, NY, USA, 383–387.

[67] Carolyn Rose and George Siemens. 2014. Shared task on prediction of dropout over time in massively open online courses. In Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs. 39–41.

[68] Belinda G. Smith. 2010. E-learning technologies: A comparative study of adult learners enrolled on blended and online campuses engaging in a virtual classroom. Ph.D. Dissertation. Capella University.

[69] Dagim Solomon. 2018. Predicting Performance and Potential Difficulties of University Student using Classification: Survey Paper. International Journal of Pure and Applied Mathematics 118, 18 (2018), 2703–2707.

[70] Denise E. Stanford-Bowers. 2008. Persistence in online classes: A study of perceptions among community college stakeholders. Journal of Online Learning and Teaching 4, 1 (2008), 37–50.

[71] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In Advances in neural information processing systems. 3104–3112.

[72] Colin Taylor ,Kalyan Veeramachaneni, and Una-May O'Reilly. 2014. Likely to stop? Predicting stop out in massive open online courses. arXiv:1408.3382

[73] Wei Wang, Han Yu, and Chuyan Miao. 2017. Deep model for dropout prediction in moocs. In Proceedings of the 2nd International Conference on Crowd Science and Engineering. ACM, ACM, New York, NY, USA, 26–32.

Bibliography (4)

[74] Pedro A. Willging and Scott D. Johnson. 2009. Factors that influence students’ decision to dropout of online courses. Journal of Asynchronous Learning Networks 13, 3 (2009), 115–127.

[75] Annika Wolff, Zdenek Zdrahal, Andriy Nikolov, and Michal Pantucek. 2013. Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In Proceedings of the third international conference on learning analytics and knowledge. ACM, ACM, New York, NY, USA, 145–149.

[76] Michalis Xenos, Christos Pierrakeas, and Panagiotis Pintelas. 2002. A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University. Computers & Education 39, 4 (2002), 361–377.

[77] Diyi Yang, Tanmay Sinha, David Adamson, and Carolyn Penstein Rosé. 2013. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In Proceedings of the 2013 NIPS Data-driven education workshop, Vol. 11. Curran Associates, Inc., USA, 14.

[78] Eran Yukseltur and Fethi Ahmet Inan. 2006. Examining the Factors Affecting Student Dropout in an Online Learning Environment. Turkish Online Journal of Distance Education 7, 3 (2006), 76–88.

[79] Zhi-Hua Zhou and Ming Li. 2005. Tri-training: Exploiting unlabeled data using three classifiers. IEEE Transactions on Knowledge & Data Engineering 17:11 (2005), 1529–1541.

[80] Mengxiao Zhu, Yoav Bergner, Yan Zhan, Ryan Baker, Yuan Wang, and Luc Paquette. 2016. Longitudinal engagement, performance, and social connectivity: a MOOC case study using exponential random graph models. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge. ACM, ACM, New York, NY, USA, 223–230.

[81] Aobakwe Senosi and George Sibiya. 2017. Classification and evaluation of privacy preserving data mining: a review. In 2017 IEEE AFRICON. IEEE, 849–855.