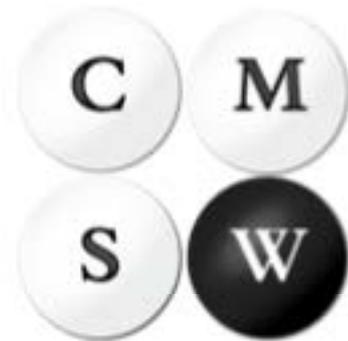# CMS.594/894- Education Technology Studio

Spring 2019

Introduction to Learning Analytics and Educational Data Mining

Class outline

- Part 1
  - Group brainstorm activity
  - Short lecture on learning analytics (LA) research
  - Mini-project #1 partner/group formation
- Part 2
  - MOOC and dataset review of the dataset
  - Learning analytics mini-project brainstorm activity
  - Share out to the class
  - Planning for next class

Brainstorm Activity

Due Today!

- You had four readings to complete today:
  - <u>Siemens & Long (2011):</u> General paper about the field of LA
  - <u>Ferguson (2012):</u> General paper about the field of LA
  - <u>Chatti et al. (2012):</u> Provides a reference model for LA based on four dimensions: what, who, why and how
  - <u>Veeramachaneni et al. (2014):</u> a technical paper that explains the feature engineering process in MOOCs
- <u>Activity:</u> Form five groups...

ACTIVITY: Brainstorming around these questions

No looking at the papers! Use your critical thinking skills to answer the following questions:

- Define your own view of learning analytics, point out your keywords
- LA has grown a lot in the last decade. What have been the enablers and drivers?
- What are the steps in the LA process? List examples of each step.
- What value can LA bring to education?
- What are the main challenges of LA?

Learning Analytics overview

Related Fields or Concepts

- **Academic Analytics**: More focused on the institutional side, e.g. use of resources or admission process
- **Action Research**: Focused on teaching practices and quality, e.g. qualitative methods, teachers and students
- **Educational Data Mining**: Developing methods, different types of ed data e.g. new Item Response Theory (IRT) algorithm or method adapted feedback
- **Recommender Systems**: Based on previous data, what items can we recommend? Content-based or Collaborative Filtering (CF)
- **Personalized Adaptive Learning**: Adaptivity (course materials adapt automatically) vs. Adaptability (learners personalize those materials)[7]

Source: Chatti et al. (2012)

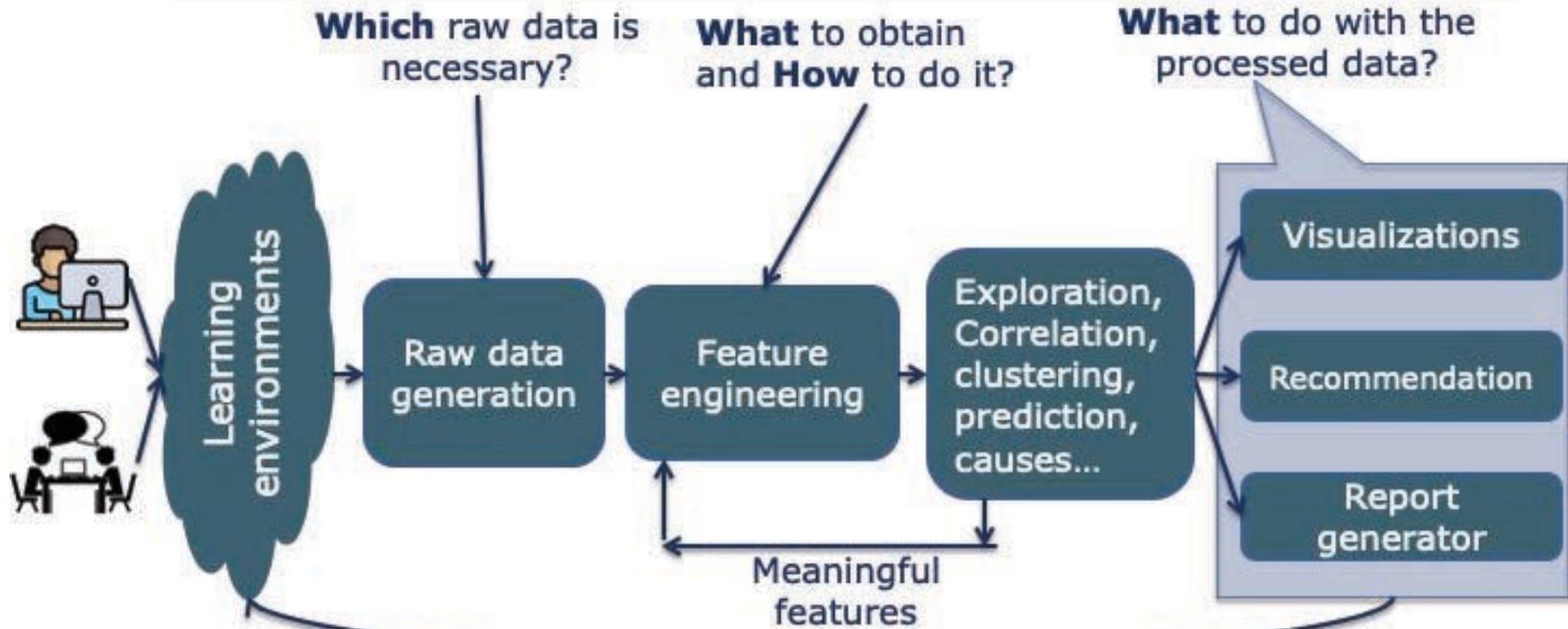# The Broad view of Learning Analytics

*…collection, analysis and reporting of **data** about <u>learners</u> and their <u>contexts</u>, for purposes of <u>understanding</u> and <u>optimising</u> **learning** and the **environments** in which it occurs…*

Source: First Learning Analytics and Knowledge Conference

LA builds on the aforementioned areas to become a generic and all-encompassing term

The Learning Analytics Process

Technology as an engine to enhance learning

Which raw data is necessary?

What to obtain and **How** to do it?

What to do with the processed data?

Learning environments — Raw data generation → Feature engineering → Exploration, Correlation, clustering, prediction, causes... → Visualizations / Recommendation / Report generator

Meaningful features

Conclusions generate feedback and close the LA loop

Reference Model of LA (Chatti et al., 2012, p.7)

We apply this reference model in the following 4 slides
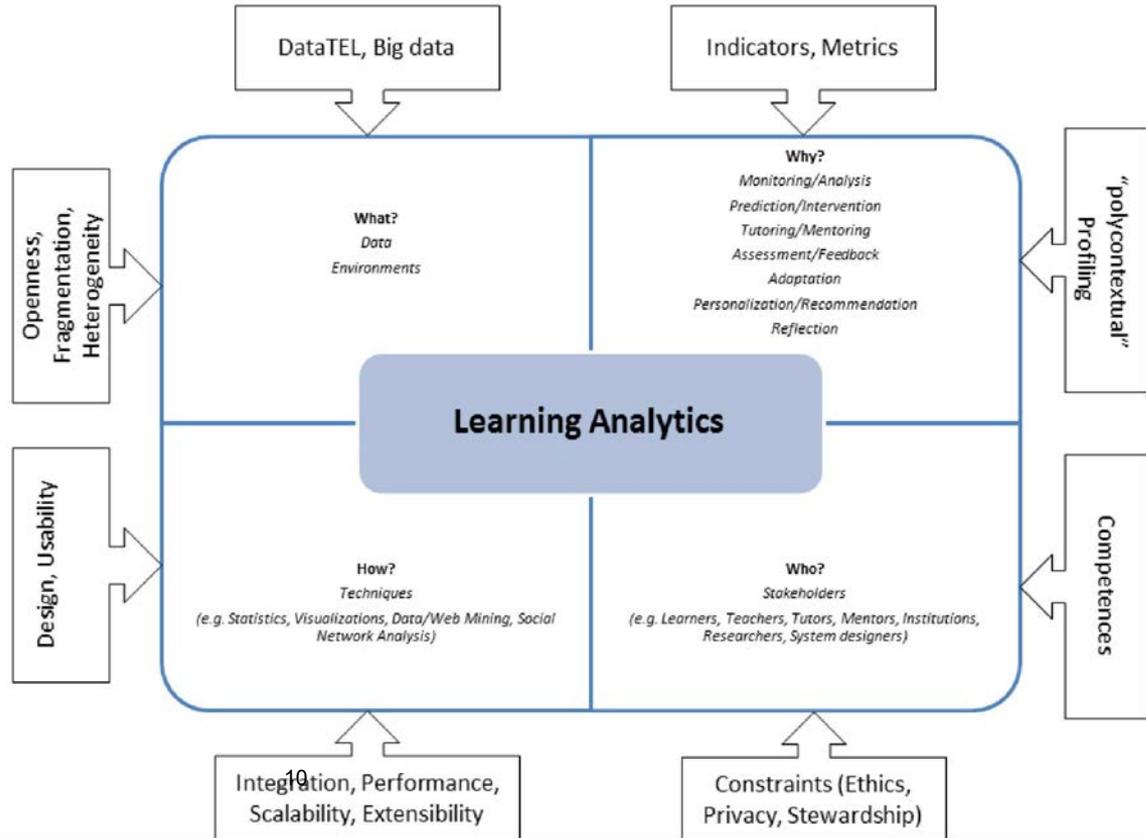
Data and Environments (What?)

What kinds of data and environments are used in the analysis?

- Learning Management Systems (Stellar,Sakai,Moodle, WebCT)
- Personal Learning Environments
- Intelligent Tutoring Systems
- Massive Open Online Courses (edX, Coursera, FutureLearn…)
- Games for Learning
- Simulation environments
- In-classroom settings using multimodal data e.g. face to face collaboration using wearables, camera and voice
- … and many more happening in informal settings! StackOverFlow, Pinterest, YouTube

Source: Chatti et al. (2012)

Stakeholders (Who?)

Who is involved?

- Students (personal data, help in learning, evaluation)
- Teachers (improve teaching, not feeling controlled, new literacy and competencies)
- Educational institutions and administrators (support decision making, students at risk)
- Private educational companies (selling their solutions)
- Governmental institutions (funding, education policy)
- Education technologists (implementation of solutions)
- Learning Analytics researchers (transfer research to practice challenge)

12

Source: Chatti et al. (2012)

Objectives (Why?)

Multiple objectives depending on the point of view of stakeholders (useful ideas or in-depth references):

- Monitoring and analysis: ANALYSE visualization dashboard
- Prediction and intervention: Self-regulation intervention
- Assessment and feedback: Immediate and adapted feedback
- Adaptation: Exercise difficulty or gaming the system
- Personalization and recommendation: Courses or threads
- In-class support: Multimodal approaches in the classroom and real-world examples in China

Source: Chatti et al. (2012)

Methods (How?)

What techniques are being applied to meet the objectives? (useful ideas or in-depth references):

- Exploratory statistics: Simple descriptive metrics, e.g. mean, mode, median, variance, charts
- Information visualization: Dashboards
- Data mining techniques:
    - Classification: Dropout prediction in MOOCs
    - Regression: Predicting learning gains
    - Clustering: Profiles of engagement in gamified environments
    - Heuristics algorithms: Cheating@Scale in MOOCs
- Social network analysis: Modeling based on graphs, e.g. interactions and participants' roles in MOOCs

14

Source: Chatti et al. (2012)

Group formation

Break - please return in 10 minutes

MOOCs and Dataset review

Massive Open Online Courses (MOOCs)

- The term MOOC was coined in 2008 after CCK08 (cMOOC)
- In 2012 Coursera, edX and Udacity emerged (xMOOC)
  - Exciting for practitioners and for human development
- Plenty of data to study the worldwide classroom
  - ...but research did not accomplish a huge impact
- A number of regional initiatives emerged (e.g. Edraak, MiriadaX, XuentangX, FUN...)
  - Many of these use Open edX software
- In 2013 Udacity announced first MOOC-based MSc
  - Seeking financial sustainability
  - Numerous Small Private Online Courses (SPOCs)

HarvardX-MITx Dataset items

- The dataset has three items
  - Creating the dataset and anonymization process:
    - EdX and ToS, raw data, edx2bigquery processing, anonymization (careful with quasi-identifiers, e.g. Netflix challenge)
  - A codebook (i.e. description the contents, structure, and layout of a dataset)
  - A comma-delimited file with the person-course data where each row represents the registration of an account to a course
- Why is this dataset important?

ACTIVITY: PLAY TIME!

MINI-PROJECT #1 BRAINSTORM

- Do you have a clear idea of what to do for your mini-project #1? If so, go for it! If not:
  - Summary of variables, draw some charts, do some readings, ask instructors...
- Design journal activity (see slide 4)
  - Idea/problem
  - Design process
  - Technical implementation
- Each group will share their slide with the class. You can work on one slide but copy to all journals
- I will share some ideas after the group shares out

More ideas – of varying technical complexity

1. Low complexity: Reports with summary statistics/charts
   - Mixed methods analysis on the influence of course in demographics and participation funnel
   - MOOC summary report for high level stakeholders i.e. donors or institutional leaders e.g. see these reports
2. Medium complexity: Viz dashboards or data mining
   - Viz dashboards: Tableau Public or R Shiny dashboard for Administrators or Instructors? E.g. see this post
   - Classification or regression models? E.g. see this study
   - Clustering of different profiles of students or courses
3. High complexity: Advanced analysis or combination
   - Combining visualization dashboards with data mining
   - Anomaly detection to find outliers, e.g. maybe to detect academic dishonesty or data processing issues
   - Network analysis of courses/students

Preparing for the next class

Due next week: first mini-project #1 analysis

- Homework
  - Additional reading on LA visualizations
  - Gain deeper knowledge of your selected method
  - Design journal assignment: Your first LA analysis or visualization

  - Need help? Let me know!
- Next week's class
  - Share out your homework, receive peer/instructor feedback and make improvements
  - In-class work on mini-projects with instructors' help to align objectives and solve tech issues
  - The work during this session will be key to your mini-project #1!

Before you leave class today…

Please complete an exit ticket

MIT OpenCourseWare
https://ocw.mit.edu

CMS.594/CMS.894 Education Technology Studio
Spring 2019