Learning analytics in (Finnish) universities - What, who, why, how?

Tommi Kärkkäinen, Mirka Saarela
Learning Analytics (LA) is an emergent discipline to "measure, collect, analyze, and report data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs." (Ferguson 2012, Siemens 2013)

Educational Data Mining (EDM) is concerned with the development of methods for exploring, understanding, and benefiting from data that come from educational settings (Romero and Ventura 2010)

Educational Knowledge Discovery (EKD): a) E-data and its environment, b) goal of study, c) tasks and methods used (incl. preprocessing and transformation), and d) knowledge obtained (Pena-Ayala 2014, Saarela and Kärkkäinen 2015)

Educational Data Science (EDS) ≈ EDM for MOOCs (Romero and Ventura 2016)

– Personal preference of general interpretation

Academic Analytics (AA) ≈ identification of meaningful patterns in educational data in order to inform and improve academic operations (Long and Siemens 2011)

Educational Analytics ≈ Knowledge discovery from educational settings (TK)


Big Data with 3-7 Vs

– Volume: size and number of observations
– Velocity: speed of data accumulation
– Variety: heterogeneity of data and formats
– Veracity: quality of data (e.g. erroneous data, nonstable distributions, missing values, . . . )
What

• Progress of studies (descriptive analytics, visualization)
  – Comparisons to
    • Plans (self)
    • Peers (others)
• Difficulties in studies (descriptive and predictive analytics, visualization)
  – Abnormal progress of studies
  – Sequencing study units in online courses
  – Supporting personal characteristics in learning – in the level and way supported by cognitive theory (cf. learning styles)
  – Other sources and forms of information
• Future of studies (predictive and prescriptive analytics, visualization)
• Student model!
Who - Stakeholders

Admin
(Organization)

Teacher

Academic advisor

Student

What I learn? When I learn? Do I learn?

Data
How – Example of Clustered Study Profiles

Credit accumulation of active students over five semesters (Autumn 2013 – 2015)

Clustered profiles of student groups
- Robust Clust
  - Missing values
  - Type of vars
- Nbr of clusters

### How – Example of Triangulated Predictive Analytics

<table>
<thead>
<tr>
<th>Course</th>
<th>3 fold CV</th>
<th>7 fold CV</th>
<th>10 fold CV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Error</td>
<td>Fleiss κ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean Grades</td>
<td></td>
<td>Mean Grades</td>
</tr>
<tr>
<td></td>
<td>6.44e-3</td>
<td>0.76</td>
<td>6.37e-3</td>
</tr>
<tr>
<td></td>
<td>1.22e-2</td>
<td>0.72</td>
<td>1.22e-2</td>
</tr>
<tr>
<td></td>
<td>R∑ rank</td>
<td>R∑ rank</td>
<td>R∑ rank</td>
</tr>
<tr>
<td>ITKP101</td>
<td>19</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>ITKP104</td>
<td>17</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>ITKA111</td>
<td>10</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>ITKA201</td>
<td>30</td>
<td>6</td>
<td>31</td>
</tr>
<tr>
<td>ITKA202</td>
<td>36</td>
<td>7</td>
<td>36</td>
</tr>
<tr>
<td>ITKA203</td>
<td>39</td>
<td>8</td>
<td>41</td>
</tr>
<tr>
<td>ITKA204</td>
<td>45</td>
<td>9</td>
<td>45</td>
</tr>
<tr>
<td>ITKP102</td>
<td>60</td>
<td>12</td>
<td>59</td>
</tr>
<tr>
<td>TIEP111</td>
<td>50</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>TIEP114</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>TIEA212</td>
<td>24</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>TIEA217</td>
<td>55</td>
<td>11</td>
<td>56</td>
</tr>
</tbody>
</table>

Saarela & Kärkkäinen: Analysing Student Performance using Sparse Data of Core Bachelor Courses. Journal of Educational Data Mining, 2015
How – Example of Dashboard

**Finland’s ranking in the PISA cycles**

**Finland’s scores in the Hofstede model dimensions**

**Characteristics of the educational system**

**Student characterization from clustering**

How – Big Data LA for Edu Syst Assmts

How – Plethodra of methods (I)

Mirka Saarela: Automatic Knowledge Discovery from Sparse and Large-Scale Educational Data - Case Finland. PhD Dissertation, University of Jyväskylä, Faculty of Information Technology, 6/2017 (to appear)
Educational Knowledge Mining

Knowledge discovery

- Goal setting
- Data selection
- Data mining
- Visual patterns and models
- Interpretation/evaluation

Data mining

- Transformation
- Method selection
- Mining
- Numerical results

- Preprocessing
- Data and tools

- Data and data sources
- Target data

- Domain analysis
- Environment

- Supervised
- Unsupervised

- Classification
- Regression
- Clustering
- Summarization
- Dependency modeling
- Link analysis
- Sequence analysis

- Exploratory Data Analysis
- Descriptive modelling
- Predictive modelling
- Discovery of patterns and rules
- Retrieval by content

- Machine learning
- Pattern recognition
- Soft computing
- Computational intelligence
- Statistical DA
- Data mining
- Business intelligence
- Knowledge discovery
- Big data analytics

(Cf. Sami Äyrämö: Knowledge mining using robust clustering. PhD Dissertation, University of Jyväskylä, Faculty of Information Technology, 2006)
The intervening factor – Who can do what for which data?

• EU’s General Data Protection Regulation (GDPR)
• The GDPR reinforces a wide range of existing rights and establishes new ones for individuals. These include
  – Right of data portability: You have the right to receive your personal data from an organisation in a commonly used form so that you can easily share it with another.
  – Right to be forgotten: you can request that an organisation delete your personal data, for instance where your data are no longer necessary for the purposes for which they were collected or where you have withdrawn your consent. (Research possibilities should be protected, though)
  – Right not to be profiled: Unless it is necessary by law or a contract, decisions affecting you cannot be made on the sole basis of automated processing.
  – (Rights to know about violations with more strict consequences)
• Purpose to encourage big data analytics of anonymised or pseudoanonymised data
• Will apply fully from 25/05/2018