Advising Analytics in Online Education

from

Capella’s Perspective

Alex Ushveridze, PhD
Agenda

• Part 1: Advising Analytics and its Deliverables
• Part 2: Scores
• Part 3: Factors
• Part 4: Dependencies
• Part 5: Interfaces
Part 1

Advising Analytics and its Deliverables
Introduction

Each educational organization has at least two mission-critical problems:

- How to increase academic performance?
- How to increase program persistence?

And there are at least two approaches to these problems:

- Department-Wide
- Learner-Centric
Department-Wide Approach

Here the starting point and further focus is department (or any sufficiently large business or academic unit).

The key steps in this approach are:

- Understand trends (on a department level)
- Find factors affecting trends (on a department level)
- Act upon factors to improve trends (on a department level)
- Through these department-level improvements, achieve overall success (learners, departments and entire organization)
Learner-Centric Approach

Here the starting point and further focus is each individual learner.

The key steps in this approach are:

- Understand trends (on an individual learner level)
- Find factors affecting trends (on an individual learner level)
- Act upon factors to improve trends (on an individual learner level)
- Through these learner-level improvements, achieve overall success (learners, departments and entire organization)
Why Take A Learner-Centric Approach?

Because:

- It is more scalable -- it can be extended to any scale through aggregation
- It is more natural and thus powerful -- if we know how to make learners more successful then we definitely know how to make the entire organization successful

Here is where advisors can really help:

- Through simply talking with learners -- this makes this approach very simple and actionable

However, here is a big challenge:

- The number of learners is so big. How can advisors frame up their communications with learners to make them maximally personalized and thus efficient?

And here is where advising analytics can help!
The goal of Advising Analytics is to help advisors optimize their conversation with learners. This includes:

- Choosing right person for conversation (Who?)
- Choosing right time for conversation (When?)
- Having clear understanding of the issue (What?)
- Having clear understanding of the causes (Why?)
- Coming up with right advice (How to?)
How Can Analytics Achieve That Goal?

Through the following four deliverables:

• Scores
• Factors
• Dependencies
• Interfaces

Let us talk about each of them separately.
1. **Scores** = numbers estimating learner future success

Scores belong to future. They answer questions: when, who, how likely?
1. **Scores** = numbers estimating learner future success
2. **Factors** = present facts responsible for learner future success

Factors belong to present. They answer questions: why? why not?

Scores belong to future. They answer questions: when, who, how likely?
Advising Analytics: Deliverables

1. **Scores** = numbers estimating learner future success
2. **Factors** = present facts responsible for learner future success
3. **Dependencies** = causal links between present and future

**Factors** (present)
- Factors belong to present.
- They answer questions: why? why not?

**Dependencies**
- Dependencies connect present and future.
- They answer questions: how to manage future from present? what if?

**Scores** (future)
- Scores belong to future.
- They answer questions: when, who, how likely?
Advising Analytics: Deliverables

1. **Scores** = numbers estimating learner future success
2. **Factors** = present facts responsible for learner future success
3. **Dependencies** = causal links between present and future

The dependencies are shown as a black box to stress the fact that they could be extremely complex and in that case not very suitable for practical use by non-technical people. There is always a tradeoff between transparency and accuracy. If we want to keep a high level of accuracy and, at the same time, let advisors use the dependencies between scores and factors in a practical way – we need to provide them with a separate unit – a specialized user interface.
1. Scores = numbers estimating learner future success
2. Factors = present facts responsible for learner future success
3. Dependencies = causal links between present and future
4. Interfaces = manageable links between present and future

Manage Future = Understand + …
Advising Analytics: Deliverables

1. **Scores** = numbers estimating learner future success
2. **Factors** = present facts responsible for learner future success
3. **Dependencies** = causal links between present and future
4. **Interfaces** = manageable links between present and future

Manage Future = **Understand** + **Act** + …
1. **Scores** = numbers estimating learner future success
2. **Factors** = present facts responsible for learner future success
3. **Dependencies** = causal links between present and future
4. **Interfaces** = manageable links between present and future

**Manage Future** = **Understand** + **Act** + **Learn from errors**
All this makes sense only if it is a continuous cyclic process including permanent improvement of analytic algorithms and advising models.
Scores
Scores may reflect different aspects of learner success and different ways of measuring it:

- **Persistence Scores:**
  - Course Registration Likelihood
  - Expected Number of Courses

- **Academic Scores:**
  - Course Completion Likelihood
  - Expected Grade
Scores: Actionability

Knowing scores means that we know whom to contact first.

This allows one to save communication resources:

<table>
<thead>
<tr>
<th>Low Persistence Score</th>
<th>Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Persistence Score</td>
<td>Do not contact</td>
</tr>
</tbody>
</table>

The more scores we know, the better.

Knowing multiple scores allows us to better segment learner populations for more targeted communications:

<table>
<thead>
<tr>
<th></th>
<th>Low Academic Score</th>
<th>High Academic Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Persistence Score</td>
<td>Communication Model 1</td>
<td>Communication Model 2</td>
</tr>
<tr>
<td>High Persistence Score</td>
<td>Communication Model 3</td>
<td>Do not contact</td>
</tr>
</tbody>
</table>
Scores: How do they look like in real life?

The simplest way to deliver scores is to create an Excel table:

This example shows tables with scores updated weekly. The scores change slightly each week (because new data is arriving) and this change is extremely informative.
Scores: How strong can their change be?

This example shows how scores may change from 2\textsuperscript{nd} to 6\textsuperscript{th} week.

Each black dot is a separate learner.

Red diagonal shows “no change”

Strong learners are above the diagonal

Weak learners are below the diagonal
Scores: What can we learn from their change?

What can we learn from these changes?

Learners at highest risk can be accurately identified at the earliest stage!

Wow! Great results! Can we replicate them?

Why this strange branch?

Oops! Something went wrong How can we help?

All these findings are highly actionable!
Part 3

Factors
Factors: Basic Types -- Examples

- Static
  - Demography

- Behavior
  - Course Registration Activity
  - Course-Room Activity
  - Post Statistics

- Content
  - Post Content
Factors: Actionability

Knowing factors means that we know what causes diverse scores to have their low or high values.

Not all factors are immediately manageable, but knowing them helps us better shape communications in hopes of improving situations in the future.

Even knowing one (dominant) factor responsible for a low score helps us select the right communication strategy (example):

<table>
<thead>
<tr>
<th>Dominant Factor</th>
<th>Optimal Communication Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low current course attendance</td>
<td># 1</td>
</tr>
<tr>
<td>Not all assignments returned</td>
<td># 2</td>
</tr>
<tr>
<td>Insufficient participation in discussions</td>
<td># 3</td>
</tr>
</tbody>
</table>
Factors: Behavior

Online Education is an ideal arena for studying learner behavior.
  • Learners have more freedom in managing their time.
  • This allows them to better expose their individuality through behavior.

Behavior patterns are well clusterable across populations.
  • People tend to behave similarly in similar situations.

Behavior patterns of individual learners are stable over time.
  • People usually do not change their habits without deep reasons.

Change of behavioral pattern is sign of something unusual.
Factors: Behavior: Example of Stability 1

Next quarter registration likelihood for currently registered learners

We can clearly see seasonal (yearly) fluctuations which demonstrate amazing stability over more than 3 years.
Next day activity likelihood for currently active learners

We can clearly see weekly fluctuations which demonstrate amazing stability over more than 3 years.
Factors: Discussion Activity -- Just Numbers

The course-room discussion activity is a very important factor of learner future success or failure.

Even very simple (numeric) variables can tell us a lot of interesting things about learner future academic performance. These are:

- Number of Posts
- Number of Threads Started
- Number of Replies
  - To Learners
  - To Instructors
- Average Reply Time
  - To Learners
  - To Instructors
Factors: Discussion Activity – Inside the Posts

The post content is another strong factor of learner future success.

| There are some very simple factors, like: | Post Length  
|                                          | Number of “I”s  
|                                          | Number of “You”s  
|                                          | Number of “?”s  
|                                          | Number of “!”s  
| We can build some composite factors, like: | Emotion = “I”s + “You”s + “!”s  
| We can create advanced factors relying on the semantic analysis of texts, like: | Text Richness  
|                                                                                                       | Independent Thinking  
|                                                                                                       | Focused-on-Subject  
|                                                                                                       | Curiosity  
|                                                                                                       | Involvement |
Factors: Posts – from Count to Content

Even simple numeric factors could be good predictors. For example, number of “I”s, “You”s or “!”s is known to be a strong negative factor of learner success, while the number of “?”s is a strong but positive factor. The number of words in a post is always a strong and positive factor.

To find stronger predictors, we need to start looking at the post content.

For example, post length is not always a good predictor because a learner’s post may contain words not written by the learner (like citations) or words unrelated to course subject. The next slides will show how to overcome this difficulty.
Factors: Posts – Original Texts

Consider a certain post (the big rectangle below) and divide it into two parts: left part – containing original texts written by learner, and right part – containing borrowed texts like templates, citations, etc. Our algorithms can do that for us.

Original text (written by learner)  Borrowed text (templates, citations)
Factors: Posts – Exploring Content

Let us now take the same post and divide it into two parts in a different way: lower part – containing course subject keywords, and upper part – containing all other words. Again, our algorithms can do that for us.
Let us now consider the overlap of the two previous divisions. We see that the lower left corner contains text which is both original and devoted to the course subject. This text characterizes learner involvement in the learning process. And the amount of this text is a very strong predictor of learner success.
Factors: Posts: Learner Involvement Example

This example shows three different views of the same text.
1. In the top view the red color marks the original text
2. In the middle view the red color marks the course subject keywords
3. And in the bottom view the red color marks the words which count for learner involvement in the learning process. The more words like this – the higher involvement – and the higher is the likelihood of learner success.
Part 4

Dependencies
(just three interesting examples)
This chart shows how learners’ next-quarter registration activity depends on the past registration pattern. The highest persistence rates have learners with more registrations in the past (with preference given to more recent registrations). This law is amazingly uniform and was actively used in our earlier predictive models.
This chart is very similar to the previous one, but instead of quarterly course registration activity we consider daily course-room activity. Again, this law is amazingly uniform and also was actively used in our earlier predictive models.
Dependencies: First weeks are a good predictors

Even very simple quantities measured before the Census Day can tell us a lot about learner further academic performance. Here we demonstrate the power of post-related quantities.
Part 5

Interfaces
Basic:

- **Real-time reporting:** Gives advisors a timely update on how each learner is doing in each of his/her courses.
- **Early warning:** Informs advisors on a timely basis about the appearance of undesirable trends/risks associated with each learner.

Advanced:

- **Coordinated action planning:** Allows advisors and instructors to share information about learners to optimize communication strategies.
- **Computer-Human Cooperation:** Allows instructors to correct computer-generated risk scores and deliver corrected scores to advisors.
ALERTS stands for the Automatic Learning Evaluation Real Time System.

This is one of the interfaces Capella was using in the past as a working prototype for more advanced proprietary tools.

The interface has several entry points. Here we show two: one for advisors and one for instructors (next two slides).

Advisor’s Interface: Advisors may select and organize desired group of learners in many different ways.

Instructor’s Interface: Instructors can see data of only those learners who are in the courserooms they currently lead.
Interface: Entry Point for Advisors

You can narrow your search. For multiple selections please use CNTRL or SHIFT keys.

<table>
<thead>
<tr>
<th>School</th>
<th>Degree</th>
<th>Type</th>
<th>Subject</th>
<th>CourseID</th>
<th>EmployeeID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any</td>
<td>Any</td>
<td>Any</td>
<td>Any</td>
<td>59400</td>
<td>Any</td>
</tr>
<tr>
<td>BTECH</td>
<td>BPA</td>
<td>Col Online</td>
<td>ART2000</td>
<td>59400</td>
<td>1000126</td>
</tr>
<tr>
<td>EDUCN</td>
<td>BS</td>
<td>Comps</td>
<td>BIO1000</td>
<td>59401</td>
<td>1000232</td>
</tr>
<tr>
<td>HASOP</td>
<td>BSN</td>
<td>Course</td>
<td>BIO1050</td>
<td>59402</td>
<td>1000277</td>
</tr>
<tr>
<td>HMSC</td>
<td>CERT</td>
<td>FirstCrisis</td>
<td>BIO2000</td>
<td>59403</td>
<td>1000310</td>
</tr>
<tr>
<td>PBSVC</td>
<td>DPA</td>
<td>Integrity</td>
<td>BPA3004</td>
<td>59404</td>
<td>1000464</td>
</tr>
<tr>
<td>UGRDS</td>
<td>EDD</td>
<td>Internship</td>
<td>BPA3035</td>
<td>59405</td>
<td>1000500</td>
</tr>
<tr>
<td></td>
<td>EDS</td>
<td>Mentor</td>
<td>BSN4000</td>
<td>59406</td>
<td>1000607</td>
</tr>
<tr>
<td></td>
<td>MBA</td>
<td>Orientatin</td>
<td>BLUS3004</td>
<td>59407</td>
<td>1000752</td>
</tr>
<tr>
<td></td>
<td>MPA</td>
<td>Practicum</td>
<td>BLUS3006</td>
<td>59408</td>
<td>1000789</td>
</tr>
<tr>
<td></td>
<td>MPH</td>
<td></td>
<td>BLUS3010</td>
<td>59409</td>
<td>1000806</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td></td>
<td>BLUS3020</td>
<td>59410</td>
<td>1000828</td>
</tr>
<tr>
<td></td>
<td>MSN</td>
<td></td>
<td>BLUS3025</td>
<td>59411</td>
<td>1000834</td>
</tr>
<tr>
<td></td>
<td>PHD</td>
<td></td>
<td>BLUS3030</td>
<td>59412</td>
<td>1000843</td>
</tr>
<tr>
<td></td>
<td>PSYD</td>
<td></td>
<td>BLUS3040</td>
<td>59413</td>
<td>1000846</td>
</tr>
</tbody>
</table>

How many records would you like to display on each page? 100 records

Pre-Sort by: Highest Risk First
## Interfaces: Entry Point for Instructors

**Alerts**
Learner performance monitoring and early warning system
Version 2.1

<table>
<thead>
<tr>
<th>Instructor Login</th>
<th>Advisor Login</th>
<th>Help</th>
</tr>
</thead>
</table>

### Instructor:

<table>
<thead>
<tr>
<th>CourseID</th>
<th>Course Start Date</th>
<th>Subject Name</th>
<th>Section</th>
<th>Category</th>
<th>Learners</th>
<th>Total Weeks</th>
<th>Completed Weeks</th>
<th>Select Course</th>
</tr>
</thead>
<tbody>
<tr>
<td>56149</td>
<td>October 5, 2009</td>
<td>PSY9919</td>
<td>44</td>
<td>Comps</td>
<td>1</td>
<td>10</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>56187</td>
<td>October 5, 2009</td>
<td>PSY8520</td>
<td>01</td>
<td>Course</td>
<td>21</td>
<td>10</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>56206</td>
<td>October 5, 2009</td>
<td>PSY7210</td>
<td>11</td>
<td>Course</td>
<td>24</td>
<td>10</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>56287</td>
<td>October 5, 2009</td>
<td>PSY8410</td>
<td>01</td>
<td>Course</td>
<td>9</td>
<td>10</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>56359</td>
<td>October 5, 2009</td>
<td>PSY7220</td>
<td>03</td>
<td>Course</td>
<td>26</td>
<td>10</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>57183</td>
<td>October 5, 2009</td>
<td>PSY7543</td>
<td>10</td>
<td>Course</td>
<td>20</td>
<td>10</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

[Open Course]
The next slides show the information pages advisors and instructors get access to after passing through the corresponding entry points.
The slide above shows the advisor’s page organized in the form of a table. Each row in the table corresponds to a separate learner in an individual course. The learner IDs and their course IDs are in the left column. All links are clickable and advisors may use them to get more information about the learners or their courses and send them or their instructors an email.

The rest of the table is organized in a way that gives advisors an instant picture of each learner’s past, present and future.

The area with blue cells describes learners’ present. The cells show how the learners are doing in their course-rooms. Each column corresponds to one week. The color of a cell is an indicator of each learner’s activity in that week. The darker cells correspond to more active learners. The gray cells mean that a learner did not show up in the course-room in that week.

The next few columns with numbers contain aggregated information about each learner’s past.

The area with green, yellow and red cells describes each learner’s future. It shows weekly dynamics of one of the learner future success scores. The green colors mark learners with higher chances to be successful, and the yellow and especially red colors mark learners with problems.

The tiny arrows in the score cells heading up or down indicate that these computer-generated scores have been corrected by instructors. We cannot exclude the situations when instructors disagree with computer scores. In that case they can correct them from their page (see next slide).
# Interfaces: Instructor’s Page

## Alerts

Learner performance monitoring and early warning system

**Version 2.1**

### Table

<table>
<thead>
<tr>
<th>CourseID</th>
<th>Course Start Date</th>
<th>Subject Name</th>
<th>Section</th>
<th>Category</th>
<th>Instructors</th>
<th>Learners</th>
<th>Total Weeks</th>
<th>Completed Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>56257</td>
<td>October 5, 2009</td>
<td>PSY8410</td>
<td>01</td>
<td>Course</td>
<td>1</td>
<td>9</td>
<td>10</td>
<td>8</td>
</tr>
</tbody>
</table>

### learner names and instructor adjustments

<table>
<thead>
<tr>
<th>EmployeeID</th>
<th>W1</th>
<th>W2</th>
<th>W3</th>
<th>W4</th>
<th>W5</th>
<th>W6</th>
<th>W7</th>
<th>W8</th>
<th>W9</th>
<th>W10</th>
<th>LQN</th>
<th>ACT</th>
<th>CPL</th>
<th>GPA</th>
<th>EAC</th>
<th>ATC</th>
<th>C_R</th>
<th>P_R</th>
<th>C_A</th>
<th>P_A</th>
<th>Correction Reason</th>
<th>Notes</th>
<th>Correction Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1053</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26</td>
<td>1.00</td>
<td>1.20</td>
<td>3.66</td>
<td>1.00</td>
<td>0.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
<td></td>
<td>2010/1/13</td>
</tr>
<tr>
<td>1069</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
<td>0.91</td>
<td>1.05</td>
<td>3.11</td>
<td>0.86</td>
<td>0.91</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>technical issues</td>
<td></td>
<td>2010/1/13</td>
</tr>
<tr>
<td>1085</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>1.00</td>
<td>1.05</td>
<td>4.00</td>
<td>1.00</td>
<td>0.79</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
<td></td>
<td>2010/1/12</td>
</tr>
<tr>
<td>1012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
<td>0.72</td>
<td>1.54</td>
<td>3.29</td>
<td>0.83</td>
<td>0.82</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>n/a</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>1031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>1.00</td>
<td>2.00</td>
<td>3.07</td>
<td>0.70</td>
<td>0.54</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td>no contact from</td>
<td></td>
<td>2010/1/13</td>
</tr>
<tr>
<td>1042</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>0.92</td>
<td>1.73</td>
<td>3.57</td>
<td>0.78</td>
<td>0.82</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td>Illness</td>
<td></td>
<td>2010/1/13</td>
</tr>
<tr>
<td>1043</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>1.00</td>
<td>1.08</td>
<td>3.42</td>
<td>1.00</td>
<td>1.00</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td>writing issues</td>
<td></td>
<td>2010/1/13</td>
</tr>
<tr>
<td>1119</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>1.00</td>
<td>1.86</td>
<td>4.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
<td></td>
<td>2010/1/13</td>
</tr>
<tr>
<td>1204</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>1.00</td>
<td>2.00</td>
<td>3.50</td>
<td>1.00</td>
<td>0.86</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td>n/a</td>
<td></td>
<td>2010/1/13</td>
</tr>
</tbody>
</table>

### Buttons

- Save Changes
- Cancel Changes
The slide above shows the instructor’s page. As before, each row in the table corresponds to a separate learner in a given course-room. The learner IDs are in the left column. The links are clickable and instructors can use them to get more information about learners and/or send them an email.

The next two (present and past) areas in this table are organized almost in the same way as in advisor’s page.

The two greenish columns show the most recent updates of two different scores reflecting the probabilities of learner future success in two different areas. As before, the green colors mark learners with higher chances to be successful, and the yellow and especially red colors mark learners with problems. The colors in these two columns are not always the same because these are different scores reflecting different aspects of a learner’s future activity.

The remaining columns allow instructors to correct computer-generated scores and leave comments explaining the reason why they disagree with computer. This portion of page essentially plays the role of communication platform between instructors and advisors. By using it they can come up with some final judgment about the learner and design the best communication strategy.
Interfaces: Computer-Human Cooperation

Nobody is perfect…

• No human can be as good as a computer in processing huge amounts of data.
• No computer can completely replace humans in making decisions.
• Both computer and human judgments could be inaccurate.
• Humans and computers should be able to correct each other

• A properly balanced interaction between computers and humans is the key to success.
Some concluding words…

• Advising can be substantially strengthened with the right analytics.

• We do not need to spend millions of dollars and hire tens of experts to have fully working advising analytics --- the solution can be astonishingly inexpensive, simple and still very powerful!

• Advising analytics can be extremely successful but only if it is organically integrated in the overall business process. This integration is not always fast and easy, and sometimes requires efforts of the entire organization. But the results are worth of these efforts!
… and one final picture

Predictive Analysts

Scores, Factors

Dependencies

Targets
Values

Who? Why? When?

What if? How to?

Interfaces

Advisors
Instructors
Leadership

Decisions
Actions

Interaction

Raw data

Learners
Thank You!

If you have questions please email me at

alex.USHVERIDZE@capella.edu