Predicting Students Success with Leganto, a proof of concept machine learning project

Mr Peter Green, Mr Gal Darom, Mr Tomer Katz, Mr David Lewis
May 2019
Leganto - Reading List Solution

• The Ex Libris Leganto software provides a Reading List solution

• Curtin Library successfully implemented in 2016

• Leganto at Curtin
  • has 2,300+ unique active courses with over 100,000 readings
  • is used over half a million times by 30,000+ unique active students
Leganto – Active courses
Leganto – Active students

Active Students

<table>
<thead>
<tr>
<th>Year</th>
<th>Active Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>2,000</td>
</tr>
<tr>
<td>2017</td>
<td>4,000</td>
</tr>
<tr>
<td>2018</td>
<td>6,000</td>
</tr>
<tr>
<td>2019</td>
<td>8,000</td>
</tr>
</tbody>
</table>

Curtin University
Proof of Concept

• Ex Libris approached Curtin with a proof of concept proposal in 2017

• Use machine learning to investigate the correlation between student success and activity within the Leganto Reading List

• Goal to have early identification of students who are likely to struggle
Benefits for Curtin

- Curtin has been using learning analytics to predict student success and to identify students at risk of failing to complete their studies.
- Project would add to early intervention strategies.
- Offer from Ex Libris to partner in a proof of concept project was accepted.
Early indication for students at risk

Advanced analytics
Gal Darom & Tomer Katz
Early indication for students at risk

\[ T = f \left( \text{Data} \right) \]

- Business Need
- Transformation
- Various type of data (structured/unstructured)
$T = f(Data)$ - What is the Business Need We Seek?

$T$ - Predicting students early in the semester that most likely will struggle with their course
\[ T = f(\textit{Data}) \] – Creating the Dataset

\textbf{Data} - Data is available from two sources:

- Curtin University
  - Students Profile
  - Students grades and academic Status
- EX Libris Leganto
  - Students engagement (usage) with the course resource list
Student Identity

- Leganto uses a hash of the student identity
- Extra data was provided with student identity hashed
- Student privacy managed
Time and Data

- Significant time and effort to produce the Curtin data
- Negotiation with Student Services and IT Services
- Leganto data needed to accumulate over multiple semesters
- Success data is reported after the event
\[ T = f (\text{Data}) \] – Creating the Dataset

• Matching Process to combine one dataset from the two data sources

• Developing meaningful features from the unified dataset to improve the model accuracy

• Examples:
  • Student engagement in relation to the average class engagement
  • Weighted student engagement per course
Complexity of Data

• People need to understand the data – conversations over the phone

• Course structure is complicated

• Student demographics complicated

• Language is not standardised – unit vs course, instructor vs coordinator

• Definition of ‘success’ matured over time
\[ T = f(\text{Data}) \] – Choosing the Algorithm

\( f \) - During the PoC several algorithm reviewed

1. Decision Tree (DT)
2. General Linear Model (GLM)
3. Naïve Baysian (NB)
4. Support vector machine (SVM)
5. Random Forest (RF)
$T = f(\text{Data})$ – Choosing the Algorithm

We found that for the relevant dataset and business need the Random Forest (RF) algorithm was the best suited classifier.
Preliminary results – Limited amount of data

The model total accuracy is 91.9%

Recall: The model will catch 18.8% of students who are at risk (25 / (108+25))

Precision: Prediction of risk student is 69.44% (For 10 students predicted as at risk, 7 will be actually at risk) (25 / (11+25))
How the Functionality May Work in the Future

Faculty / Semester

New Semester

Identify students at risk

Improve Model

Prediction model

Leganto usage data

Curtin data
Thank you!

Gal.Darom@exlibrisgroup.com
Tomer.Katz@exlibrisgroup.com
The Future

• Proof of Concept is ongoing as more time and data can improve the model

• Too early for Ex Libris to know if this might feature in Leganto in the future
Thank you

Make tomorrow better.