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

Mrs Linda Sheedy
August 2019
With thanks to

Mr Peter Green, Mr Gal Darom, Mr Tomer Katz, and Mr David Lewis

For their contributions to this project and this presentation
Curtin University

Learning and Teaching being delivered from

Perth Western Australia
(including Bentley, and Perth City)
Kalgoorlie Western Australia
Malaysia Campus
Mauritius Campus
Singapore Campus
Dubai
Leganto - Reading List Solution

• In August 2015, Curtin Library chooses the Ex Libris Leganto as its new Reading List solution.

• Curtin Library successfully implemented in the first Semester 2017.

• Between 2016 and August 2019, teaching staff have:
  • created 4,700+ Reading Lists (includes Reading Lists migrated from previous system)
  • added over 115,300 citations which resulted in
  • 1.5 million (non-unique) full text views of resources by 42,400+ unique active students.
Ex Libris approached Curtin with a proof of concept proposal in 2017 to use machine learning to investigate the correlation between student success and activity within the Leganto Reading List. 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

\[ T = f \left( Data \right) \]

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

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

**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
Time and Data

- Significant time and effort to produce the Curtin data
- Negotiation with Student Services and Digital and Technology Solutions
- 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)
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

Leganto usage data
Curtin data

Prediction model
Identify students at risk

New Semester
Improve Model
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!
Gal.Darom@exlibrisgroup.com
Tomer.Katz@exlibrisgroup.com

Make tomorrow better.
Images

A large blank world map

https://upload.wikimedia.org/wikipedia/commons/c/cf/A_large_blank_world_map_with_oceans_marked_in_blue.PNG (This file is licensed under the Creative Commons Attribution-Share Alike 2.5 Generic, 2.0 Generic and 1.0 Generic license.)