Building a machine learning model to predict the polarity of student feedback comments

Richard Arnett, Joanna Zawadzka & Catriona Keane. RCSI Quality Enhancement Office

- *"I found this lecture to be taught well. However, there was some confusion related to the coursework assignment and what was required."*
- "♥ ズロ◆■≏ ◆ ごそ・ M M ◆ M M ◆ □ & M

 少 ジョニ ◆ ・ M ● ④ □ M ◆ M □ ⊕ ◆ ご M □ M ⑤ •

 □ M M □ ズ ◆ H □ □ M ⑤ ◆ M ≏ ◆ □ ◆ ご M

 M □ ◆ □ M □ 0 & S • + Y ○ M ◆ ⑤ ≏ ご ⑥ ◆ • •
- ~20,000 (non-blank) comments 2014-2016
- ~40,000 sentences
- ~3.2 million words



Methodology

Process: classify; split sentences; lower case; remove punctuation; remove common words

1. found lecture taught well (Positive)

2. however confusion related coursework assignment required (Negative)

| A | Variables | | | | | | | 0 | utcome | | |
|----------|-----------|---------|--------|------|---------|-----------|---------|------------|------------|----------|-------|
| Sentence | found | lecture | taught | well | however | confusion | related | coursework | assignment | required | Class |
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | Pos |
| 2 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | Neg |

Split Data: Training 75%: Test 25%

Train models on Training Data

Test models on Test Data

Model complexity: Logistic Regression (simple) -> Random Forests (complex)

Metric: Classification 'Accuracy'

Target: 'No Information Rate' (64%)



Results

| Model | Time (secs) | Accuracy (%) |
|---------------------------------------|--------------|--------------|
| Logistic Regression | 18.01 | 84% |
| Linear Discriminant Analysis (LDA) | 20.36 | 83% |
| Elastic Net Regression | <u>67.17</u> | <u>85%</u> |
| Support Vector Machine (SVM) | 650.5 | 83% |
| Random Forest | 7236.5 | 82% |



| | 0 | Occurence (%) | | | |
|--|-------------|---------------|--|--|--|
| Comments | Positive | Negative | | | |
| I found this lecture to be taught well | <u>0.62</u> | 0.38 | | | |
| However, there was some confusion related to the coursework assignment and what was required | 0.25 | <u>0.75</u> | | | |



В

D