Lost in the Woods: Creating a Predictive Typology of Those Who Wander off the Path

Terrence Willett, Steve Blohm, & Megan Leonard
Cabrillo College | 2018 RP Group
Leakage

Terrence Willett, Steve Blohm, & Megan Leonard
Cabrillo College | 2018 RP Group
Leakage

Why do students drop out?

**Short Term - Fall to Spring Persistence**

Why do we lose students from fall to spring? Who are we losing?

**Long Term - Enrollment Pattern Analysis**

What enrollment patterns are present in cohorts? Who completes their education goal? Who drop out permanently?
Short Term: Fall to Spring Persistence
Fall to Spring Persistence

Returning In The Spring After Enrolling in The Fall

- 69.35% Enrolled in Spring
- 30.65% Leaked

All students from 2010-11 to 2016-17
7% to 10% received degrees in the fall, but many came back. Less than 5% left mid-year after receiving a degree. An improving economy can lead to loss of enrollments. At Cabrillo we lost many part-time students.
Loss of Part Time Students in Fall Semesters

We are losing more part time students than full time students as the years go by.
Loss of Part Time Students in Fall Semesters

We are losing more part time students than full time students as the years go by.
Fall to Spring Persistence- Methodology

Pulled data from Fall 2010 to Spring 2017. Examined persistence from Fall semesters to the following Spring semesters

- Fall 2010 - Spring 2011
- Fall 2011 - Spring 2012
- ...
- Fall 2016 - Spring 2017

Shoutout to Terra Morris! Terra built the original non-persister query that this query was based on.
Fall to Spring Persistence - Variable Selection

Predictor Variables:
- Number of Fall credits
- GPA
- Cumulative units attempted
- Age
- URM
- Gender
- Economically disadvantaged
- EOPS
- Used Disability Services
- Veteran
- Foster Youth
- AB540
- Transfer education goal

Outcome Variable:
- Enrollment in subsequent Spring semester (binary variable)
Fall to Spring Persistence – Enrollment Status

Overall: 87% of full time students persisted from fall to spring while only 60% of part time students persisted from fall to spring.

<table>
<thead>
<tr>
<th></th>
<th>Fall 2010</th>
<th>Fall 2011</th>
<th>Fall 2012</th>
<th>Fall 2013</th>
<th>Fall 2014</th>
<th>Fall 2015</th>
<th>Fall 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Time</td>
<td>86.13%</td>
<td>87.23%</td>
<td>86.08%</td>
<td>86.78%</td>
<td>86.48%</td>
<td>86.83%</td>
<td>87.31%</td>
</tr>
<tr>
<td>Part Time</td>
<td>61.76%</td>
<td>63.72%</td>
<td>58.92%</td>
<td>58.49%</td>
<td>61.34%</td>
<td>58.61%</td>
<td>60.19%</td>
</tr>
</tbody>
</table>
Students with an online enrollment persist to spring at a slightly higher rate of 72.4% (vs. 69.4%). But students with only online enrollments persist at a rate of 46%. 
Descriptive Statistics - Initial Math Placement

Spring Enrolled

- Did not Assess: 61.28%
- Retest: 69.84%
- Basic Math: 71.66%
- Elementary Algebra: 72.52%
- Intermediate Algebra: 76.39%
- Transfer Math: 80.63%
Descriptive Statistics - Past Math Enrollments

Re-Enrollment in Spring Semester by Whether Student Had Taken Math at Cabrillo College

- Not Taken Math: 53.51% enrolled in spring, 46.49% no enrollment
- Taken Math: 76.69% enrolled in spring, 23.31% no enrollment
Descriptive Statistics - Past English Enrollments

Re-Enrollment in Spring Semester by English Enrollment Behavior

- Not Taken English:
  - Enrolled in Spring: 53.45%
  - No Enrollment: 46.55%

- Taken English:
  - Enrolled in Spring: 76.19%
  - No Enrollment: 23.81%
Descriptive Statistics - Education Goal

Re-Enrollment in Spring Semester by Education Goal

- No Transfer Goal: 63.59% Enrolled in Spring, 36.41% No Enrollment
- Transfer Ed Goal: 75.30% Enrolled in Spring, 24.70% No Enrollment
Descriptive Statistics - Financial Aid

Re-Enrollment in Spring Semester by Financial Aid Use

- No Financial Aid: 65.09%
- Uses Financial Aid: 82.24%

- Enrolled in Spring: 34.91%
- No Enrollment: 17.76%
Fall to Spring Persistent - Logistic Regression

A logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable.

Independent Variables: The Kitchen Sink

Outcome Variable: Persistence to Spring
Logistic Regression

Many analyses later...

<table>
<thead>
<tr>
<th>Step 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>0.286</td>
<td>0.01</td>
<td>2175.04</td>
<td>1</td>
<td>0</td>
<td>1.331</td>
</tr>
<tr>
<td>Full-Time Student</td>
<td>1.245</td>
<td>0.02</td>
<td>4235.44</td>
<td>1</td>
<td>0</td>
<td>3.474</td>
</tr>
<tr>
<td>Uses DSPS</td>
<td>0.737</td>
<td>0.03</td>
<td>860.627</td>
<td>1</td>
<td>0</td>
<td>2.09</td>
</tr>
<tr>
<td>Taken Math</td>
<td>0.699</td>
<td>0.02</td>
<td>1995.88</td>
<td>1</td>
<td>0</td>
<td>2.012</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.82</td>
<td>0.02</td>
<td>1630.67</td>
<td>1</td>
<td>0</td>
<td>0.443</td>
</tr>
</tbody>
</table>
Fall to Spring Persistence - Logistic Regression

Results

- The more units someone takes in the Fall, the more likely they are to re-enroll in the Spring
- Taking Math and English are good predictors
- Students who receive accessibility benefits are often more likely to e-enroll in Spring.
Predicting Persistence - ANOVA

Fall to Spring Persistence

- FallCumulativeUnits $\geq 18$
  - FallCumulativeUnits $\geq 5.5$
    - 0.12 (4%)
  - FallCumulativeUnits $\geq 44$
    - Age $< 20$
      - 0.59 (14%)
    - 0.68 (31%)
  - FallCumulativeUnits $\geq 50$
    - 0.76 (6%)
    - 0.83 (8%)

- FallCumulativeUnits $\geq 18$
  - FallEnrollmentStatusAbbr = FT
    - 0.38 (11%)
    - 0.93 (26%)

cp = 0.005; Cross Validated Error = 0.80604
Predicting Persistence - Poisson

- \( cp = 0.005 \); Cross Validated Error = 0.86541

### Fall to Spring Persistence

- **FallCumulativeUnits \( \geq 18 \)**: 0.89 (32%)
  - **FallCumulativeUnits \( \geq 4.5 \)**: 0.67 (53%)
    - **FallEnrollmentStatusAbbr = PT**: 0.088 (1%)
    - **FallEnrollmentStatusAbbr = FT**: 0.39 (10%)
    - **FallEnrollmentStatusAbbr \( \neq \) PT**: 0.1 (3%)
- **FallCumulativeUnits \( < 18 \)**: 0.1 (3%)
  - **FallEnrollmentStatusAbbr = PT**: 0.005 (0.1%)
  - **FallEnrollmentStatusAbbr = FT**: 0.088 (1%)

---

### Model Parameters

- Cross-validated error: 0.86541
- CP: 0.005

---

### Notes

- The model uses a Poisson distribution to predict persistence from Fall to Spring.
- Key features include cumulative units and enrollment status.
Predicting Persistence - Classification

Fall to Spring Persistence

\[ \text{FallCumulativeUnits} \geq 18 \]

- No: 0 (15%)
- Yes: 1 (85%

cp = 0.005; Cross Validated Error = 0.81517
Predicting Persistence - Classification

Fall to Spring Persistence

- FallCumulativeUnits ≥ 18
  - yes: 1 (85%)
  - FallCumulativeUnits ≥ 5.5
    - no: 0 (4%)
    - yes: 0 (1%)
      - FallEnrollmentStatusAbbr = PT
        - DisabilityEver = DSP
          - 0 (9%)
          - 1 (1%)

\( cp = 0.0025; \) Cross Validated Error = 0.80698
library(rpart)
library(rpart.plot)

# set control parameters
ctrl005 <- rpart.control(minsplit = 100, cp = 0.005, xval = 10)

# send formula to object
f1 <- SpringEnrolled ~ Age + URM + Gender + EconomicallyDisadvantaged + EOPSEver + VeteranStatus + DisabilityEver + FosterYouthStatus + FirstGenerationStatus + LocalZip + AB540 + EducationGoal + EnglPlacement + MathPlacement + FallTerm + FallCumulativeUnits + FallEnrollmentStatusAbbr

# reference formula object in rpart command
t1 <- rpart(formula = f1, data = p, method = "anova", control = ctrl005)

# tree output
printcp(t1)
print(t1)
rsq.rpart(t1)

prp(t1, main = "Fall to Spring Persistence", extra = 100, varlen = 0, left = FALSE, box.palette = "RdYlGn")
Long Term: Enrollment Pattern Analysis
Enrollment Pattern Analysis - The Origin

Crosta (2013). Intensity and attachment: How the chaotic enrollment patterns of community college students affect educational outcomes.
Enrollment Pattern Analysis - The Origin

Crosta (2013). Intensity and attachment: How the chaotic enrollment patterns of community college students affect educational outcomes.

Examined the relationship between enrollment patterns and outcomes

- Enrollment Intensity (full time vs part time)
- Enrollment Continuity (enrolling consecutively vs skipping one or more terms)
Enrollment Pattern Analysis - The Origin

Crosta (2013). Intensity and attachment: How the chaotic enrollment patterns of community college students affect educational outcomes.

Examined the relationship between enrollment patterns and outcomes

- Enrollment Intensity (full time vs part time)
- Enrollment Continuity (enrolling consecutively vs skipping one or more terms)

Positive relationships between

- Intensity & likeliness of transferring
- Continuity & likeliness of graduating with a certificate or degree
Enrollment Pattern Analysis - Methodology

Gathered six years of data (12 primary semesters + 6 summer semesters) Fall 2011 - Spring 2017 from Cabrillo’s SQL Data Warehouse.

2100 lines of SQL code; 280+ data elements; many hours of data cleaning

Analyzed enrollment patterns using a k-means cluster analysis using SPSS.

Used enrollment pattern clusters alongside other variables to map out the likelihood that students would graduate and/or transfer.

- Decision tree (SPSS)
Enrollment Pattern Analysis - Sankey Diagram

Enrollment Pattern Sankey Diagram

- Full Time
- Part Time
- Not Enrolled
- Graduate
- Transfer
- Dropped
Enrollment Pattern Analysis – Pattern Creation

Enrollment Patterns

- 0 for part time enrollment
- 1 for full time enrollment
- . for no enrollment
- Concatenated all and gathered data based on enrollment pattern
  - Enrollment switches such as from PT - FT and FT - PT (both taking into account gaps and not)
  - Consecutive terms with similar behavior such as PT - PT and FT - FT
  - First time interruptions of one, two, and three consecutive semesters
Enrollment Pattern Analysis - Common Patterns

Common Enrollment Patterns Overall (n = 2702)

- 0............ (enrolled one term part time, then dropped out/completed) n = 531
- 00.......... (enrolled one year part time, then dropped out/completed) n = 173
- 11.......... (enrolled one year full time, then dropped out/completed) n = 106
- 1............ (enrolled one term full time, then dropped out/completed) n = 86
Enrollment Pattern Analysis - Common Patterns

Common Enrollment Patterns Overall (n = 2702)

- 0…………… (enrolled one term part time, then dropped out/completed) n = 531
- 00………… (enrolled one year part time, then dropped out/completed) n = 173
- 11………… (enrolled one year full time, then dropped out/completed) n = 106
- 1…………… (enrolled one term full time, then dropped out/completed) n = 86

Common Enrollment Patterns of Known Completions (n = 790, 29.24%)

- 0…………… (enrolled one term part time, then completed) n = 96
- 00………… (enrolled one year part time, then completed) n = 31
- 11………… (enrolled one year full time, then completed) n = 42
- 0…………… (enrolled one term full time, then completed) n = 30
Enrollment Pattern Analysis - Common Patterns

Common Enrollment Patterns of Those Who Graduate (n = 423, 15.66%)

- 11111111... (enrolled three years full time, then graduated) n = 37
- 11111111... (enrolled two years full time, then graduated) n = 16
- 111111111... (enrolled four years full time, then graduated) n = 13
- 111111110... (enrolled five consecutive terms full time, then one term part time, then graduated) n = 4
Enrollment Pattern Analysis - Common Patterns

Common Enrollment Patterns of Those Who Transfer (n = 431, 15.95%)

- 0........ (enrolled one term part time, then transferred) n = 96 *
- 00......... (enrolled one year part time, then transferred) n = 31
- 11......... (enrolled one year full time, then transferred) n = 26
- 1111........ (enrolled two years full time, then transferred) n = 17

* likely these students were not true first time students, may have been four-year students/etc
Enrollment Pattern Analysis - Cluster Analysis

K-Means Cluster Analysis

- Aims to partition observations into a certain number of clusters with each observation belonging to the cluster with the nearest mean, serving as a prototype of the cluster.

We ran a K-Means Cluster Analysis to create enrollment pattern clusters.

Variables were those gathered about the enrollment patterns.
Enrollment Pattern Analysis – Cluster Analysis

K-Means Cluster Analysis - Variables

- Total number of terms enrolled
- Total number of switches in enrollment (full to part time, part time to not enrolled, etc)
- Full time to part time switches, ignoring any gaps in enrollment
- Full time to full time transitions, ignoring any gaps in enrollment
- Part time to part time transitions, ignoring any gaps in enrollment
- Part time to full time switches, ignoring any gaps in enrollment
Enrollment Pattern Analysis – Cluster Analysis

K-Means Cluster Analysis – Even More Variables

- Consecutive terms not enrolled
- Full time to part time consecutive switches
- Full time to full time consecutive transitions
- Part time to part time consecutive transitions
- Part time to full time consecutive switches
- First single term interruption
- First two-consecutive-terms interruption
- First three-consecutive-terms interruption
Enrollment Pattern Analysis – Clusters

- Cluster 0: Continuous Switchers
- Cluster 1: Short Term Leavers
- Cluster 2: Full Time Persisters
- Cluster 3: Longer Term Leavers
- Cluster 4: Revolving Doors
- Cluster 5: Returners
- Cluster 6: Part TimePersisters
<table>
<thead>
<tr>
<th>Cluster</th>
<th>Size</th>
<th>Description</th>
<th>Sample</th>
<th>Completion %</th>
</tr>
</thead>
</table>
| 0. Continuous Switchers | 233  | Analysis rejects were primarily individuals who enrolled consecutively in all terms | 001010110110001011010000 | Graduated: 34.29%  
|                      |      |                                                                             |                         | Transferred: 8.93%  |
| 1. Short Term Leavers | 1027 | Enrolled one or two semesters and then left permanently                      | 0...........01........... | Graduated: 0.88%    
|                      |      |                                                                             |                         | Transferred: 18.79% |
| 2. Full TimePersisters| 290  | Enrolled consecutively for 6+ terms, mostly FT, some switches between FT and PT | 111111......00111100.... | Graduated: 61.72%   
|                      |      |                                                                             |                         | Transferred: 15.52% |
| 3. Longer Term Leavers| 451  | Enrolled mostly PT for a few terms and then left permanently                 | 000...........010........ | Graduated: 9.98%    
|                      |      |                                                                             |                         | Transferred: 17.96% |
| 4. Revolving Doors   | 280  | Enrolled and then left several times for short periods of time, many switches in enrollment | 0111...1....00.111.0.... | Graduated: 15.71%   
|                      |      |                                                                             |                         | Transferred: 16.79% |
| 5. Returners         | 182  | Enrolled and then left for a longer period of time (often 2+ years) before returning | 0.....0000..10......010.| Graduated: 7.14%    
|                      |      |                                                                             |                         | Transferred: 11.54% |
| 6. Part Time Persisters | 125 | Enrolled consecutively PT for many terms, only left for one semester at a time, if at all | 000.0000....010000.00... | Graduated: 11.20%   
<p>|                      |      |                                                                             |                         | Transferred: 10.40% |</p>
<table>
<thead>
<tr>
<th></th>
<th>Continuous Switchers</th>
<th>Short Term Leavers</th>
<th>Full TimePersisters</th>
<th>Longer Term Leavers</th>
<th>Revolving Doors</th>
<th>Returners</th>
<th>Part TimePersisters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female %</td>
<td>52.16%</td>
<td>46.93%</td>
<td>52.41%</td>
<td>49.22%</td>
<td>45.71%</td>
<td>47.25%</td>
<td>56.80%</td>
</tr>
<tr>
<td>Avg Age at Start</td>
<td>20.75 years</td>
<td>26.14 years</td>
<td>19.31 years</td>
<td>22.93 years</td>
<td>20.51 years</td>
<td>22.13 years</td>
<td>22.45 years</td>
</tr>
<tr>
<td>% &lt; 26</td>
<td>88.18%</td>
<td>68.35%</td>
<td>93.10%</td>
<td>80.27%</td>
<td>89.29%</td>
<td>78.02%</td>
<td>85.60%</td>
</tr>
<tr>
<td>% White</td>
<td>46.69%</td>
<td>52.87%</td>
<td>46.55%</td>
<td>50.78%</td>
<td>49.64%</td>
<td>43.96%</td>
<td>40.00%</td>
</tr>
<tr>
<td>% LatinX</td>
<td>44.38%</td>
<td>33.30%</td>
<td>41.72%</td>
<td>33.48%</td>
<td>38.21%</td>
<td>40.11%</td>
<td>45.60%</td>
</tr>
<tr>
<td>% URM</td>
<td>51.01%</td>
<td>41.67%</td>
<td>50.00%</td>
<td>43.68%</td>
<td>46.79%</td>
<td>49.45%</td>
<td><strong>55.20%</strong></td>
</tr>
<tr>
<td>% Local</td>
<td><strong>88.18%</strong></td>
<td>66.21%</td>
<td>77.93%</td>
<td>73.61%</td>
<td>76.43%</td>
<td>84.62%</td>
<td>81.60%</td>
</tr>
<tr>
<td>% DSPS</td>
<td>19.02%</td>
<td>3.99%</td>
<td>7.24%</td>
<td>7.98%</td>
<td>6.79%</td>
<td>8.24%</td>
<td><strong>24.80%</strong></td>
</tr>
<tr>
<td>% Pell</td>
<td>73.78%</td>
<td>40.41%</td>
<td>72.76%</td>
<td>65.41%</td>
<td>62.86%</td>
<td>60.44%</td>
<td>56.80%</td>
</tr>
<tr>
<td>% Transfer Ed Goal</td>
<td>64.84%</td>
<td>39.34%</td>
<td>68.97%</td>
<td>58.31%</td>
<td>61.79%</td>
<td>46.15%</td>
<td>43.20%</td>
</tr>
</tbody>
</table>
Enrollment Pattern Analysis – Decision Tree

- Decision tree analyses are non-parametric approaches to modeling data that result in a series of if-then “rules”
- The decision trees for MMAP used binary splits based on the Gini-Simpson index
- CHAID = Chi-Square Automatic Interaction Detector (Kass 1980)
  - Allows multi-nominal splits
  - Uses Chi-square distribution to determine splitting criteria
Full Time Persisters

Node 4

<table>
<thead>
<tr>
<th>Category</th>
<th>%</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>26.9</td>
<td>78</td>
</tr>
<tr>
<td>1.000</td>
<td>73.1</td>
<td>212</td>
</tr>
<tr>
<td>Total</td>
<td>10.7</td>
<td>290</td>
</tr>
</tbody>
</table>

URM

Adj. P-value=0.017, Chi-square=5.682, df=1

URM

Node 11

<table>
<thead>
<tr>
<th>Category</th>
<th>%</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>33.1</td>
<td>48</td>
</tr>
<tr>
<td>1.000</td>
<td>66.9</td>
<td>97</td>
</tr>
<tr>
<td>Total</td>
<td>5.4</td>
<td>145</td>
</tr>
</tbody>
</table>

Node 12

<table>
<thead>
<tr>
<th>Category</th>
<th>%</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>20.7</td>
<td>30</td>
</tr>
<tr>
<td>1.000</td>
<td>79.3</td>
<td>115</td>
</tr>
<tr>
<td>Total</td>
<td>5.4</td>
<td>145</td>
</tr>
</tbody>
</table>
Enrollment Pattern Analysis - Results

Full Time Persisters are most likely to graduate and/or transfer

But there is disproportionate impact for URM students

Attempting transfer level Math or English in their first term increases the likelihood that they transfer and/or graduate
Future Steps

Fall to Spring Persistence

- Fall to Fall persistence
- Students who are not enrolling in the first place

Enrollment Pattern Analysis

- Use enrollment patterns to predict outcomes (logistic regression or similar)
- Add more cohorts, see if enrollment patterns are changing over time
- Include summer terms

Suggestions?
Questions?