LEVERAGING PREDICTIVE ANALYTICS TO DRIVE STUDENT ACCESS & SUCCESS

Presented by,
Harshitha Akula
PREDICTIVE ANALYTICS (BACKGROUND)

- University had an existing external predictive analytics provider
- Report users wanted to better predict student success
- Administrators wanted to better allocate staff time & resources
- IER wanted to build predictive models in-house
LOCALLY DEVELOPED

- No dependency on external providers
- Flexible & Cost efficient solution
- Leveraged existing university reporting software
- Partnership between IER & ICBE
WEBFOCUS R-STAT

- Add-on to existing university reporting software
- Leverages the power of R
- Graphical User Interface (GUI)
- Ease of predictive model deployment to existing user reports
- Better use of staff time & resources
- Limitless applications
BUSINESS OBJECTIVE

Increase freshmen enrollment for fall 2018

IER Solution:

• Create enrollment probabilities for all admitted freshmen students
• Deploy enrollment model to existing Enrollment Management reports
• Rank students by enrollment probabilities to focus on top enrollment targets

INTERNAL USER EXAMPLE

• Enrollment Management Division:
  • VP of Enrollment Management
  • Admissions Director
  • Financial Aid
  • Registrar office
  • Admissions Recruiters
The model splits a data mining project into six phases and it allows for needing to go back and forth between different stages.
WebFOCUS R-STAT delivers powerful predictive analytics functionality. Business users can leverage a single integrated solution for BI, data modeling, and scoring, so they can make decisions based on accurate, validated future predictions instead of relying on gut instinct alone.
### STEPS FOR R-STAT:

![R-STAT Interface](image)

**Page 1**

<table>
<thead>
<tr>
<th>ID</th>
<th>COLL_CODE1</th>
<th>DEGC_CODE1</th>
<th>MAJR_CODE1</th>
<th>HIGH_SCHOOL_PERCENTILE</th>
<th>GENDER</th>
<th>FINAID_APPLICANT_IND</th>
<th>ADMIT</th>
<th>ENROLLED</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axxxx</td>
<td>Axxxx</td>
<td>Axxxx</td>
<td>Axxxx</td>
<td>111,111,111,111,&gt;</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>Axxxx</td>
</tr>
<tr>
<td>Bxxxx</td>
<td>Bxxxx</td>
<td>Bxxxx</td>
<td>Bxxxx</td>
<td>222,222,222,222,&gt;</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>Bxxxx</td>
</tr>
<tr>
<td>Axxxx</td>
<td>Axxxx</td>
<td>Axxxx</td>
<td>Axxxx</td>
<td>111,111,111,111,&gt;</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>Axxxx</td>
</tr>
<tr>
<td>Bxxxx</td>
<td>Bxxxx</td>
<td>Bxxxx</td>
<td>Bxxxx</td>
<td>222,222,222,222,&gt;</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>Bxxxx</td>
</tr>
</tbody>
</table>
Roles noted. 3000 observations and 10 input variables. The target is ENROLLED. Categoric 2. Classification models enabled.

<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Data Type</th>
<th>Input</th>
<th>Target</th>
<th>Risk</th>
<th>Ident</th>
<th>Ignore</th>
<th>Weight</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ID</td>
<td>Numeric</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>Unique: 3000</td>
</tr>
<tr>
<td>2</td>
<td>COLL_CODE1</td>
<td>Categoric</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>Unique: 5</td>
</tr>
<tr>
<td>3</td>
<td>DEGC_CODE1</td>
<td>Categoric</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>Unique: 13</td>
</tr>
<tr>
<td>4</td>
<td>MAJR_CODE1</td>
<td>Categoric</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>Unique: 57</td>
</tr>
<tr>
<td>5</td>
<td>HIGH_SCHOOL_PERCENTILE</td>
<td>Numeric</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>Unique: 99 Missing: 1</td>
</tr>
<tr>
<td>6</td>
<td>GENDER</td>
<td>Categoric</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>Unique: 2</td>
</tr>
<tr>
<td>7</td>
<td>FINAID_APPLICANT_IND</td>
<td>Categoric</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>Unique: 2</td>
</tr>
<tr>
<td>8</td>
<td>ADMIT</td>
<td>Constant</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>Unique: 1</td>
</tr>
<tr>
<td>9</td>
<td>ENROLLED</td>
<td>Categoric</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>Unique: 2</td>
</tr>
<tr>
<td>10</td>
<td>ALL_RACE</td>
<td>Categoric</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>Unique: 9</td>
</tr>
<tr>
<td>11</td>
<td>AGE</td>
<td>Numeric</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>Unique: 21</td>
</tr>
<tr>
<td>Variable</td>
<td>Levels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEGC_CODE1</td>
<td>13 integer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAJR_CODE1</td>
<td>57 integer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIGH_SCHOOL_PERCENTILE</td>
<td>double 123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENDER</td>
<td>2 integer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FINAID_APPLICANT_IND</td>
<td>2 integer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADMIT</td>
<td>1 integer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALL_RACE</td>
<td>9 integer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>integer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENROLLED</td>
<td>2 integer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Summary statistics generated.
Summary of the Ada Boost model:

Call:
ads(ENROLLED ~ ., data = crs8dataset(crs8train, c(crs8input, crs8target)), pars = list(split = "information"), bag.frac = 1, control = rpart.control(maxdepth = 30, cp = 0.01, minsplit = 20, xval = 10, usesurrogate = 0, maxsurrogate = 0), iter = 50)

Loss: exponential, Method: discrete, Iteration: 50

Final Confusion Matrix for Data:

<table>
<thead>
<tr>
<th>True Value</th>
<th>Final Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1199</td>
</tr>
<tr>
<td>N</td>
<td>153</td>
</tr>
<tr>
<td>Y</td>
<td>288</td>
</tr>
<tr>
<td>Y</td>
<td>460</td>
</tr>
</tbody>
</table>

Train Error: 0.21

Out-of-Bag Error: 0 iteration= 6

Additional Estimates of number of iterations:
In Excel:
Any Questions ?