# Predicting Students' Course Success Using Machine Learning Approach

Akira Kanatsu Research Analyst Tanner Carollo Assistant Director

Amanda Bain Graduate Student Research Assistant

CAIR 2018 Annual Conference - November 14, 2018





### Outline

CSUSB WE DEFINE THE Future

- At a Glance
- Why Course Success
- Machine Learning Approach
- Data and Model Training
- Results
- Next Steps



## At a Glance

- Undergraduate Student Population
  - 17,854 (89%)

Fall 2018 Lecture/Seminar Undergraduate Courses (< 500 level)

- 656 courses
- 1,380 sections
- 58,370 total enrollments
- 2,800 (5%) of enrollments are repeats
- 8,023 attempted enrollments (courses full)



# Why Course Success



# Impact on Retention, Graduation, and Y2D





## Impact at the Course Level

- 2017-18 Lower Division LEC/SEM Courses
  - 55 (16%) had a DFWI rate of 20% or higher
  - 10,418 of the course enrollments were repeats
  - 8,012 attempted enrollments received a course is full message





# Machine Learning Approach



# Machine Learning

#### A branch of artificial intelligence

- Learn from data
- Identify patterns
- Make decisions

#### Application Examples

- Online recommendations
- Self-driving cars
- Fraud detection



# Application

**IBM SPSS Modeler** (2018 CSU Innovation Minigrant)

- Powerful data mining software
- Accepts various data sources
- Visual interface without programing Language
- Automated modeling/model training
- Ability to apply trained models to a separate data









### **Courses Examined**



CSUSB WE DEFINE THE Future

2326

SSCI

564

PHIL

## **Course Success Predictors**

#### Student Information

- General
- Term Specific
- Course Information
  - Instructor Information

- Demographics (e.g., Sex & URM)
- Student Level
- Basis of Admission
- Degree Type (e.g., BA)
- College of Major
- Incoming GPA (i.e., HS & TR GPA)
- SAT & ACT Scores
- Developmental Math/Placement



## **Course Success Predictors**

#### Student Information

- General
- Term Specific
- Course Information
  - Instructor Information

- Cumulative GPA
- Term Attempted Units
- Course Repeat/Highest Previous Grade
- # of High DFWI Courses (>25%)
- Average DFWI Rates
- # of Concurrent Courses with 10% Higher DFWI Rates When Taken Together



## DFWI Rates by Math Course Sections



## **Course Success Predictors**

#### Student Information

- General
- Term Specific
- Course Information
- Instructor Information

- Term (e.g., Fall)
- Campus (SB & Palm Desert)
- Meeting Days (e.g., MWF)
- Class Begin Time
- Instruction Mode (e.g., Online)

Class Size



### **Course Success Predictors**

#### Student Information

- General
- Term Specific
- Course Information
- Instructor Information

Tenure Status
Previous Course Teaching Experience
Average GPA
Average DFWI Rates



# Model Training



|  | s <u>W</u> indow <u>H</u> elp                            |                          |                                 |               |              |
|--|--|--------------------------|---------------------------------|---------------|--------------|
| 7 🚍 🖬 🗛 🗴 🗞 🖻 👝 🗠  | 🗄 📝 🛑 🍉 🔶 🎸  | : 👍 💽 🛯                  | 246<br>246                      |               |              |
|  |  |                          |                                 |               |              |
|  |  |                          |                                 |               |              |
|  | 🎯 Туре   |                          |                                 |               | ×            |
|  | Proview  |                          |                                 |               |              |
|  | Lieview  |                          |                                 |               |              |
|  | $\geq$   |                          |                                 |               |              |
|  | Types Format Apportations                                |                          |                                 |               |              |
|  |  |                          |                                 |               |              |
|  | Read   | Tvalues Clear var        | des Clear Air Values            |               |              |
|  | Field -  | Measurement              | Values                          | Missing Check | Role         |
|  | A CAD_YK_LAB   | Flag<br>Nominal          | 2017.187/2016.17"               | None          | None         |
|  | TEST GRP   | Nominal                  | 1.0,2.0,3.0                     | None          | ○ None       |
|  | D_YYT_LAB  | 💑 Nominal                | "Fall 2016","Fall 2017","S      | None          | None None    |
|  | TERM_LAB   | 💑 Nominal                | Fall,Spring,Summer,Winter       | None          | > Input      |
|  | D_CRS_LAB  | Flag                     | MATH110/MATH110                 | None          | ○ None       |
| BIOL200_2156toLastTe Type                                  | A CLASS SECTION  | Nominal                  | "01" "02" "03" "04" "05" "0     | None          | O None       |
|  | A CAMPUS   | Flag                     | PALM/MAIN                       | None          | None<br>None |
|  | A DAYS   | 💑 Nominal                | MTWR,MW,MWF,TR                  | None          | 🔪 Input      |
|  | A B_TIME   | 💑 Nominal                | "0100PM","0120PM","020          | None          | N Input      |
|  |  | Nominal                  | "0110PM","0150PM","023          | None          | None None    |
|  |  | Continuous               | [21 0 48 0]                     | None          | None Input   |
|  | FACULTY_ID   | Continuous               | [1579.75,1448768.25]            | None          | ○ None       |
|  | TENURE   | 🎖 Flag                   | YES/NO                          | None          | 🔪 Input      |
|  | SEC_TAUGHT_PREV  | Continuous               | [1.0,9.0]                       | None          | N Input      |
|  | AVC. CPA   | Continuous               | [20.0,350.0]                    | None          | ○ None       |
|  | DEWL COUNT   | Continuous               | [0.64,3.29]                     | None          | S input      |
|  | Ø DFWI_RATE  | Continuous               | [4.55,65.22]                    | None          | > Input      |
|  | STUDENT_ID   | n Continuous             | [672113.25,1675455.25]          | None          | ○ None       |
|  | CRS_CNT_PREV   | Continuous               | [0.0,4.0]                       | None          | Input        |
|  | HIGHEST_GRADE_PREV                                       | Continuous               | [6.0,28.0]                      | None          | None Input   |
|  | A GRADE  | Nominal                  | A.AB.B+.BC.C+.CBE I             | None          | Some         |
|  | ADJUSTED_GRADEN  | Continuous               | [0.0,4.0]                       | None          | Ø None _     |
|  |  | A Nominal                | 0.010                           | Nono          | Torgot 🔳     |
|  | View current fields O View u                             | unused field settings    |                                 |               |              |
|  |  |                          |                                 |               |              |
|  | UK Cancel  |                          |                                 |               | Apply Reset  |
|  |  |                          |                                 |               |              |
|  |  |                          |                                 |               |              |
|  |  |                          |                                 | ***           |              |
| Favorites Sources Record Ops Field Ops 🔺 Grag              | hs 🔷 Modeling 🔳 Output 🔳 Export 🎯 IBM® SPS               | S® Statistics 🖳 Pyth     | ion 🏂 Spark                     |               |              |
|  |  | • (PDD) (-•              |                                 |               |              |
|  |  |                          |                                 |               |              |
| o Data Prep Type Filter Derive Filler Reclassify Anonymize | Binning RFM Analysis Ensemble Partition SetToFlag Restru | ucture Transpose History | Field Reorder Reproject Time In | ervals        |              |
|  |  |                          |                                 |               |              |







### Results



# Model Output

| Prediction – whether students are       |
|---|
| likely to receive DFWI (1=Yes, 0=No)    |
| Confidence – how likely these           |
| predictions are correct (range: 0 to 1) |

|             | Prediction | Confidence  |
|-------------|------------|-------------|
| Student 155 | 1          | 0.77777778  |
| Student 156 | 1          | 0.666666667 |
| Student 157 | 1          | 0.666666667 |
| Student 158 | 1          | 0.88888889  |
| Student 159 | 1          | 0.88888889  |
| Student 160 | 1          | 0.77777778  |
| Student 161 | 1          | 0.666666667 |
| Student 162 | 0          | 0.77777778  |
| Student 163 | 0          | 0.77777778  |
| Student 164 | 0          | 0.77777778  |



# Model Prediction & Accuracy

|        |            | Historical (Summer 2015 to Summer 2018) |                       |                    |                    |  | Fall 2018  |                       |                       |
|--------|------------|---|-----------------------|--------------------|--------------------|--|------------|-----------------------|-----------------------|
| Course | Enrollment | Predicted<br>DFWI (n)                   | Predicted<br>DFWI (%) | Actual<br>DFWI (n) | Actual<br>DFWI (%) | Overall Model Accuracy<br>on Historical Data | Enrollment | Predicted<br>DFWI (n) | Predicted<br>DFWI (%) |
| ADMN   | 2709       | 598                                     | 22%                   | 641                | 24%                | 80%  | 133        | 39                    | 29%                   |
| BIOL   | 1276       | 629                                     | 49%                   | 645                | 51%                | 77%  | 408        | 175                   | 43%                   |
| MATH   | 8363       | 2287                                    | 27%                   | 2136               | 26%                | 81%  | 681        | 102                   | 15%                   |
| PHIL   | 653        | 267                                     | 41%                   | 241                | 37%                | 79%  | 103        | 34                    | 33%                   |
| SSCI   | 3557       | 1004                                    | 28%                   | 1231               | 35%                | 74%  | 407        | 84                    | 21%                   |



# Top 5 Predictors (CHAID Model)

| Rank | ADMN                 | BIOL                    | MATH                 | PHIL                   | SSCI             |
|------|----------------------|-------------------------|----------------------|------------------------|------------------|
| 1    | Cumulative GPA       | Cumulative GPA          | Instructor DFWI Rate | Cumulative GPA         | Cumulative GPA   |
| 2    | Instructor DFWI Rate | Student Level           | Cumulative GPA       | Instructor Average GPA | Tenure Status    |
| 3    | Tenure Status        | Dev. Math Tenure Status | Dev. Math Grade      | Days                   | College of Major |
| 4    | Class Size           | SAT Score               | Incoming GPA         | Concurrent CRS         | Term             |
| 5    | Term                 | URM Status              | Concurrent CRS       | Sex                    | SAT Score        |



## MATH Course Prediction by Section

|         | <b>DFWI Prediction</b> |     | Total      | Predicted        | Predicted DFWI Rate with | Instructor      |
|---------|------------------------|-----|------------|------------------|--------------------------|-----------------|
| Section | Yes                    | No  | Enrollment | <b>DFWI Rate</b> | > 60% Expected Accuracy  | Historical DFWI |
| 01      | 6                      | 34  | 40         | 15%              | 5%                       | 11%             |
| 02      | 2                      | 38  | 40         | 5%               | 0%                       | 9%              |
| 03      | 1                      | 39  | 40         | 3%               | 3%                       | 16%             |
| 04      | 35                     | 6   | 41         | <b>85%</b>       | 76%                      | 43%             |
| 05      |                        | 40  | 40         | 0%               | 0%                       | 9%              |
| 06      |                        | 40  | 40         | 0%               | 0%                       | 0%              |
| 07      | 4                      | 36  | 40         | 10%              | 5%                       | 19%             |
| 08      | 1                      | 40  | 41         | 2%               | 0%                       | 5%              |
| 09      | 1                      | 39  | 40         | 3%               | 0%                       | 5%              |
| 10      | 29                     | 11  | 40         | 73%              | 38%                      | 43%             |
| 11      | 2                      | 38  | 40         | 5%               | 3%                       | 16%             |
| 12      |                        | 40  | 40         | 0%               | 0%                       | 15%             |
| 13      | 4                      | 36  | 40         | 10%              | 8%                       | 15%             |
| 14      | 15                     | 25  | 40         | 38%              | 20%                      | 34%             |
| 15      | 2                      | 38  | 40         | 5%               | 3%                       | 16%             |
| 80      |                        | 39  | 39         | 0%               | 0%                       | 8%              |
| 81      |                        | 40  | 40         | 0%               | 0%                       | 8%              |
| Total   | 102                    | 579 | 681        | 15%              | 9%                       | 16%             |

# Other Possible Report: MATH Student with Concurrent Course

|             | Identified Concurrent Course | Number of Concurrent Courses |
|-------------|------------------------------|------------------------------|
| Student 234 | Yes                          | 1                            |
| Student 241 | Yes                          | 2                            |
| Student 267 | Yes                          | 1                            |
| Student 277 | Yes                          | 1                            |
| Student 278 | Yes                          | 1                            |
| Student 279 | Yes                          | 1                            |
| Student 280 | Yes                          | 1                            |
| Student 284 | Yes                          | 1                            |
| Student 287 | Yes                          | 1                            |
| Student 299 | Yes                          | 2                            |

### Next Steps



#### Process Improvement

- Exploration and Automation

#### Targeted Supplemental Instruction

#### Advising

- Notify students where/when appropriate (e.g., concurrent courses)

#### Reporting outputs and structures







### PLEASE FILL OUT AN EVALUATION FOR THIS SESSION

Rate now >

| November 14, 2:45 PM - 4:30 PM |
|--------------------------------|
| Catalina C3                    |
| 大 古 古 古 古 No Ratings           |
| DESCRIPTION                    |

#### CLICK on the EVALUATION LINK > IN YOUR CAIR APP

THANK

Evaluate this session

Predictive analytics are being utilized more and more frequently in higher education as we aim to determine ways we can better determine which students are likely to be successful on our campus. With holistic data becoming more readily available and advanced statistical techniques becoming more higher-education friendly, it's clear that innovative uses of data are not merely some passing fad. Yet, for campus stakeholders, figuring out ways to start making use of data and conducting predictive analyses can be a daunting task. In this bootcamp, we will work with live <u>datasets together to</u> determine

### YOU!

