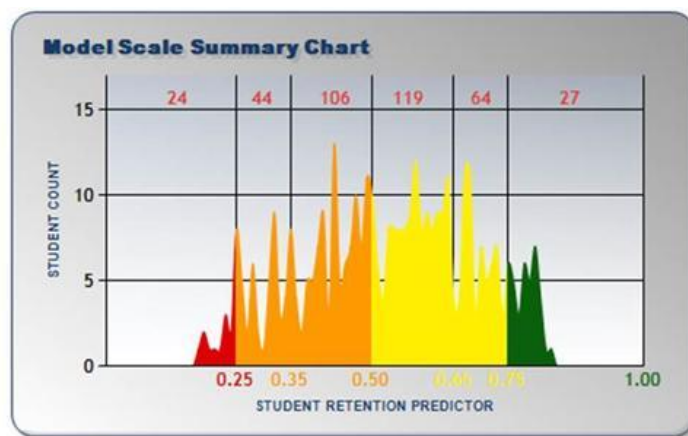**Predictive Analytics in Higher Education**
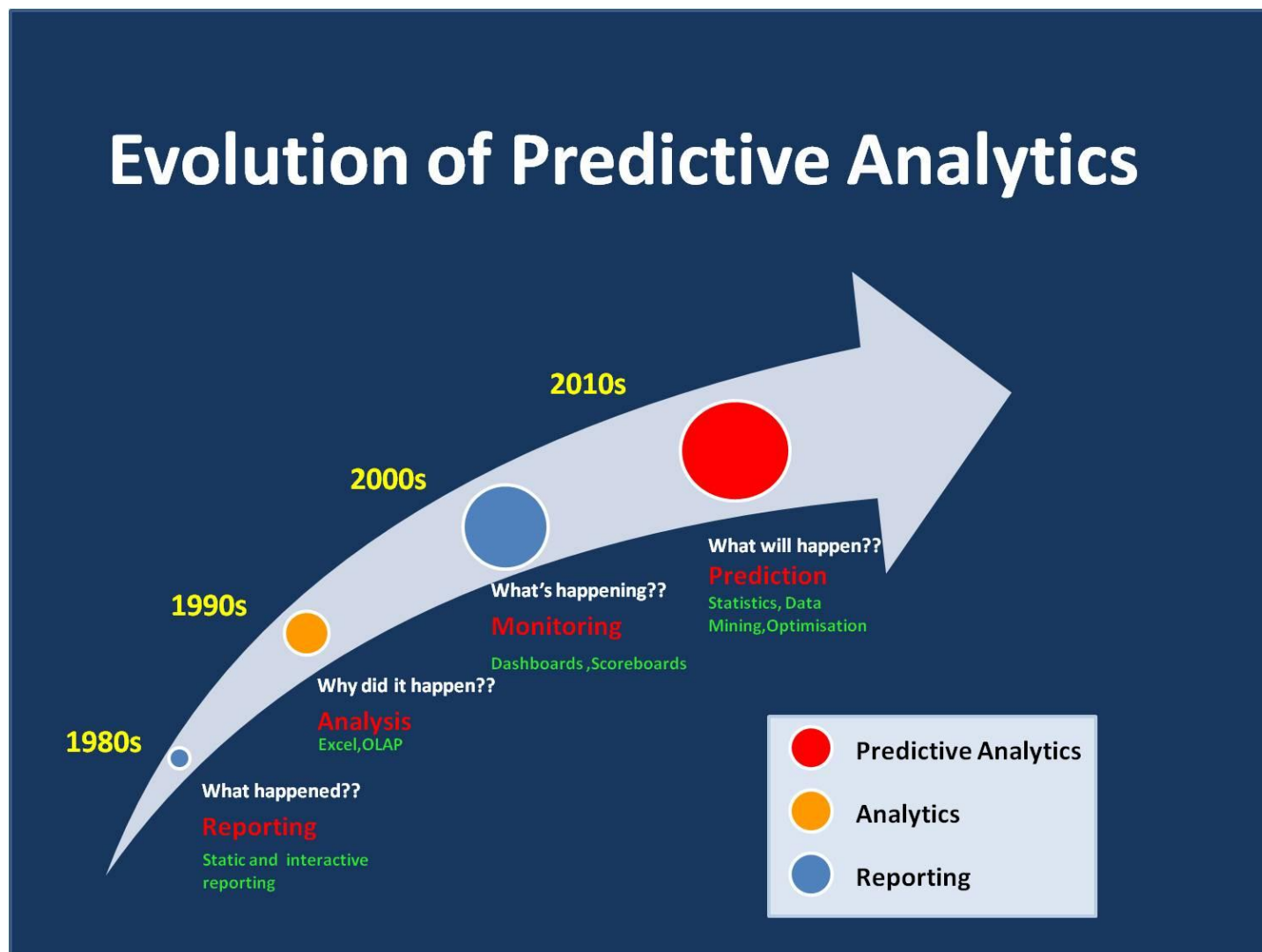(And an OCCC Trial Demonstration)

Presentation By:

Matthew Eastwood
Research Analyst
Office of Institutional Effectiveness

April 8th, 2015

How Has Data Analysis in Higher Education Changed Over the Years?

What is Predictive Analytics you ask?

Predictive analytics is the practice of extracting information from existing data sets in order to determine patterns and predict future outcomes and trends. Predictive analytics does not tell you what will happen in the future. It forecasts what might happen in the future with an acceptable level of reliability.

Source: http://www.webopedia.com/TERM/P/predictive_analytics.html

What have been some traditional industry sources utilizing predictive analytics?

- Actuarial Science
- Financial Services
- Insurance
- Manufacturing
- Telecommunications
- Retail
- Travel/Logistics
- Marketing
- Energy/Utilities
- Airlines

More recent large industry sources:
- Healthcare
- Education
- Government

Currently, Predictive Analytics and Data Mining is 'En Vogue' throughout the Business Professional Communities

Who's getting hired and what are they doing?"

To get to an answer, we analyzed the skills and experience data in over 330 million LinkedIn member profiles. If your skills fit one of the categories below, there's a good chance you either started a new job or garnered the interest of a recruiter in the past year.

Source: http://blog.linkedin.com/2014/12/17/the-25-hottest-skills-that-got-people-hired-in-2014/

The 25 Hottest Skills of 2014 on LinkedIn

1. Statistical Analysis and Data Mining
2. Middleware and Integration Software
3. Storage Systems and Management
4. Network and Information Security
5. SEO/SEM Marketing
6. Business Intelligence
7. Mobile Development
8. Web Architecture and Development Framework
9. Algorithm Design
10. Perl/Python/Ruby
11. Data Engineering and Data Warehousing
12. Marketing Campaign Management
13. Mac, Linux and Unix Systems
14. User Interface Design
15. Recruiting
16. Digital and Online Marketing
17. Computer Graphics and Animation
18. Economics
19. Java Development
20. Channel Marketing
21. SAP ERP Systems
22. Integrated Circuit (IC) Design
23. Shell Scripting Languages
24. Game Development
25. Virtualization

The Early Adopters
How have Higher Education Institutions Utilized Predictive Analytics?

1. Increasing Success/Retention Rates

Example: Georgia State University

Outcome: Increased graduation rates by 22 points

How: The university analyzed some 2 million historic grades and modeled how performance in one class might predict performance later on. Data showed that, not surprisingly, students needed to do well in foundational courses for their chosen major to avoid struggling in higher-level classes. Supplemental Instruction (paid high achieving student tutors) was subsequently targeted to classes like Introduction to Accounting.

Source: http://www.nationaljournal.com/next-economy/solutions-bank/the-intuitive-but-innovative-secret-behind-1-university-s-turnaround-20130923

How have Higher Education Institutions Utilized Predictive Analytics?

2. Targeting Prospective Students for Admission

Example: Creighton University

Outcome: Recruitment Cost Savings of $28,000

How: Used predictive modeling to reach those most likely to enroll without communicating to the entire population of prospective students. They were able to eliminate 35,000 of their lowest scoring prospects from their mailing efforts.

How have Higher Education Institutions Utilized Predictive Analytics?

3. Identify underprepared students when they enter campus
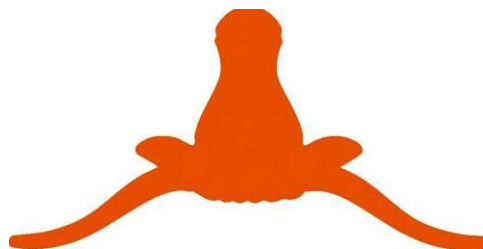
Example: University of Texas - Austin

Outcome: Intervention group had good standing GPA (>= 2.0) for 87% of students compared to 78% for like students not provided the intervention

How: The University Leadership Network (ULN), started in the 2013–2014 academic year, is a scholarship and experiential learning program which uses predictive analytics to identify students with academic and financial need and help them develop leadership skills. Students receive $5,000 per year over four years, paid out in ten $500 payments. During their first-year, students in ULN work on academic and professional development while performing community service, participating in discussion groups, and attending weekly seminars on topics such as time management, professional branding, and team building. In their second-year, students intern in various on-campus offices where they learn real-world job skills.

Source: http://www.theuia.org/sites/default/files/UIA_predictive_onepagers.pdf

How have Higher Education Institutions Utilized Predictive Analytics?

4. Helping students identify majors related to their interests and mapping courses to achieve their degree

Example: Arizona State University
Outcome: The 2009 cohort has already achieved a four-year graduation rate that is 12 percentage points higher than before eAdvisor was introduced.

How: Developing eAdvisor to help undergraduates identify majors related to their interests has helped reduce the number of exploratory majors at the university from one-third of freshman students to only 8 percent of freshman. Students not only find majors through the system, but they map their classes and track progress toward completing their degrees. Launched for the 2008 entering class, the system has been expanded to all undergraduate students.

Source: http://www.changemag.org/Archives/Back%20Issues/2013/January-February%202013/improving-advising-full.html

Currently, Information Tech. Companies are Marketing Heavily towards Higher Education

A Classic Example of Predictive Analytics:
How Auto Insurance Companies Predict Risk to Base their Rates

What are the at risk groups within the variables?

1. Driving History – Prior Wrecks/Citations/Suspensions
2. Age – Under 25
3. Gender – Males
4. Marital Status – Single
5. Students – Below a B Average
6. Credit Scores – Poor Credit Rating/Debt/No Mortgage
7. Occupation – Doctors/Lawyers/Business Executives
    (Lowest Risk = First Responders and Teachers!)
8. Location – High Crime Rates/Heavy Traffic Neighborhoods/Rural
9. Vehicle Type – High Theft Rates/Model (Sports Car)
10. Weather – Severe Weather Areas
11. Terrain – Mountains/Coastal
12. Several More….

Source: http://www.dmv.org/insurance/auto-insurance-rates-101.php

An OCCC *Example* Demonstration

Question: Based on information we know, can we accurately predict which new to OCCC students will pass all classes and which students will fail one or more courses within the first week of fall classes?

(Much like the auto insurance example...which students are at the highest risk of failure?)

Fall 2013 New Student (No Concurrent) Actual Outcomes

Failed One or More 45%

Passed All 55%
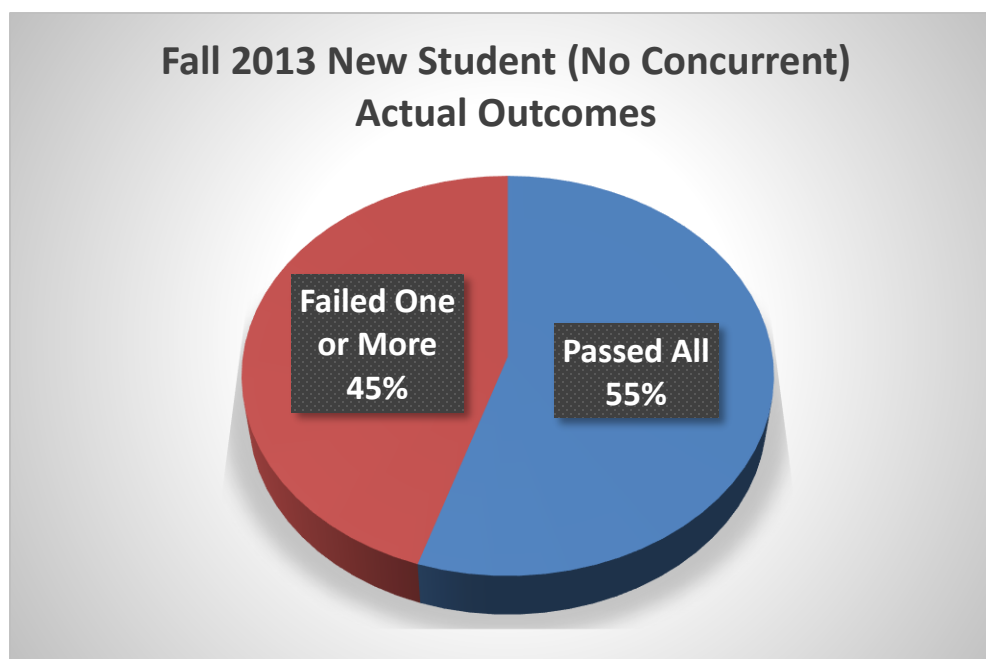
Looking Back at Fall 2013 to Predict Fall 2014

Using information on hand regarding students fall of 2013 at the beginning of the semester, which variables impacted on eventual grade predictor model.

The 20 Variables Used

Online Enrolled (One or More)

Evening Enrolled (One or More)
Class Days of the Week (One or More 1 day a week class)
Course Duration (One or More non-16 week - No Developmental)
Math Enrolled - 1513 or Higher
SCL Enrolled
Developmental Enrolled
Ethnicity
Gender
Age
Zip Code
Education Goals (Type of Degree Sought, Plan on Transfer, etc...)
New or New Transfer
High School

High School Graduation Type
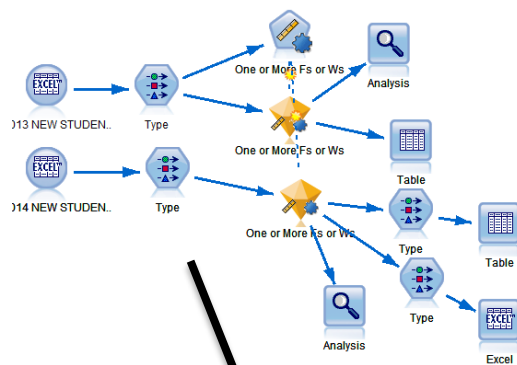ACT Scores

Declared Major at Admissions
Household Income
Financial Need
Father Education

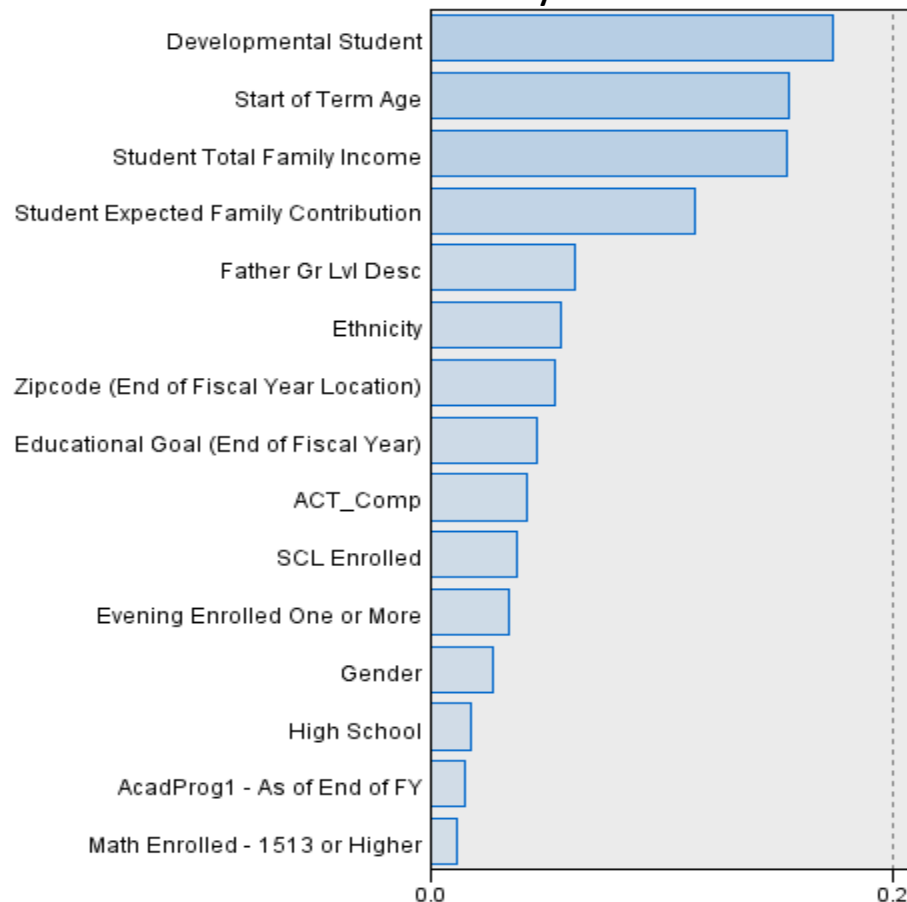# Inputting Fall 2013 Grades as the Dependent Variable

Process the Models…

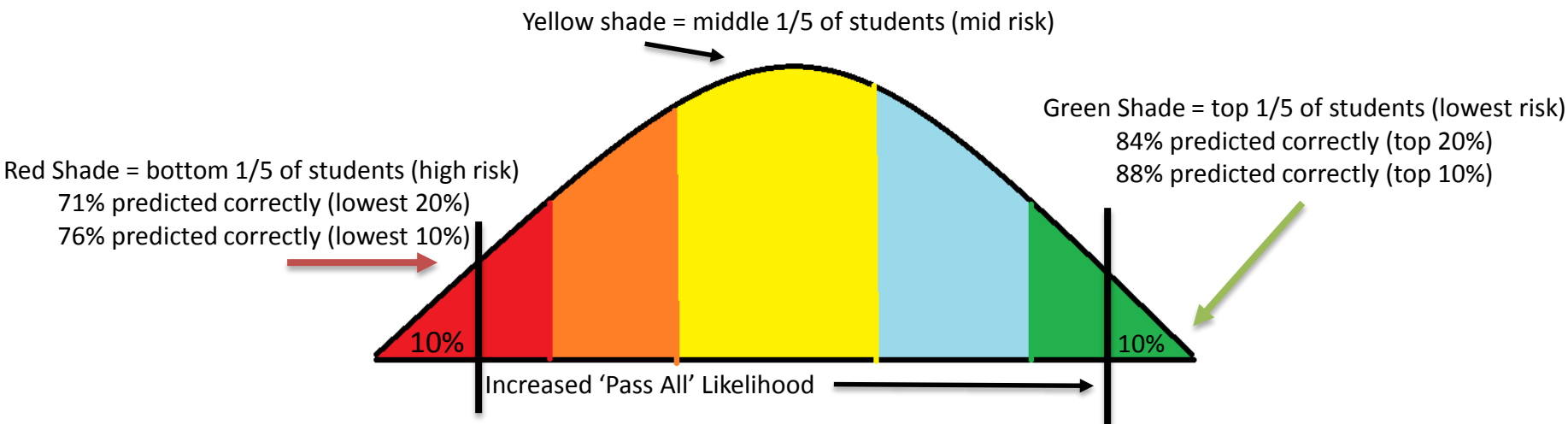
Select Which Models to Include…

Use?	Graph	Model
✔		CHAID 1
✔		Generalized Linear 1
✔		Linear 1

Analyze the Results…

Developmental Student
Start of Term Age
Student Total Family Income
Student Expected Family Contribution
Father Gr Lvl Desc
Ethnicity
Zipcode (End of Fiscal Year Location)
Educational Goal (End of Fiscal Year)
ACT_Comp
SCL Enrolled
Evening Enrolled One or More
Gender
High School
AcadProg1 - As of End of FY
Math Enrolled - 1513 or Higher

0.0   0.2

*GED, Online (One or More), One Day a Week (One or More), Course Duration, and New/New Transfer Status had no predictor importance in the model.

14

Applying the Fall 2013 Model to Fall 2014 (using the same variables from prior slide) to predict if a student would pass all or fail one or more courses by the end of the semester.

Yellow shade = middle 1/5 of students (mid risk)

Green Shade = top 1/5 of students (lowest risk)
84% predicted correctly (top 20%)
88% predicted correctly (top 10%)

Red Shade = bottom 1/5 of students (high risk)
71% predicted correctly (lowest 20%)
76% predicted correctly (lowest 10%)

10%   10%

Increased 'Pass All' Likelihood

| Model Outputs	Headcount	Percents
Predicted Fail - Highest Risk Group (top 20%)	**739**	**20.0%**
Actual Fail One or More	521	70.5%
Actual Pass All	218	29.5%
Predicted Pass All - Lowest Risk Group (bottom 20%)	**739**	**20.0%**
Actual Fail One or More	122	16.5%
Actual Pass All	617	83.5%
Unknown	**2,216**	**60.0%**
Actual Fail One or More	1,023	46.2%
Actual Pass All	1,193	53.8%
Grand Total	**3,694**	**100.0%**

And Finally, How Is This Applied to Individual Students…
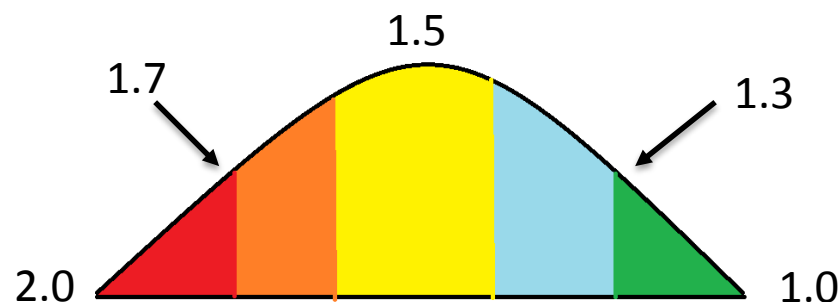
Prediction Confidence Codes
2.0 = Most Confident will fail
1.0 = Most Confident will pass
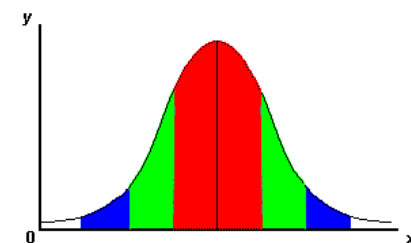
Fake Name	Devl Student	SCL Enrolled	Gender	Start of Term Age	Educational Goal	ACT	Student Total Family Income	Father Gr Lvl	Actual Pass or Fail?	Predicted Output	Prediction Confidence
Joffrey	Yes	No	M	19	02	16	$29,552	Jr. High	Actual Fail	Predicted Fail	1.88
Jon S	No	No	M	19	02	21	$125,781	High School	Actual Fail	Unknown	1.56
Brienne	No	No	F	28	07				Actual Pass	Predicted Pass	1.18
Tyrion	Yes	Yes	M	31	01				Actual Pass	Unknown	1.54
Ygritte	Yes	Yes	F	19	01	16	$23,506	Unknown	Actual Fail	Predicted Fail	1.72
Hodor	No	No	M	44					Actual Pass	Predicted Pass	1.15
Grey Worm	Yes	Yes	M	19	01		$49,087	High School	Actual Pass	Predicted Fail	1.72

Conclusion: Predictive analytics will by no means act as a definitive answer to what a student will experience, but it can offer insights into likely outcomes if all things remain the same.