

The background of the entire slide is a dark blue field filled with a complex network of thin, light-colored lines connecting numerous small, glowing orange-yellow dots, creating a sense of a global or digital network.

A MODEL FOR SUCCESS

Using Machine Learning and Data Analytics to
Determine Incoming Students that will Require
Additional Support to Achieve Future Success

DEFINING SUCCESS

Job Satisfaction (F3B34D) and How your job aligns with your career goals (F3B29).

Please indicate the extent to which you agree or disagree with each of the following statements about your [current/most recent job]:

You feel fairly well satisfied with your present job.

1=Strongly agree

2=Agree

3=Neither agree nor disagree

4=Disagree

5=Strongly disagree

Which of the following best describes your [current/most recent job]?

1=It fulfills your long-term career goals.

2=It is a step on the path to your long-term career goals.

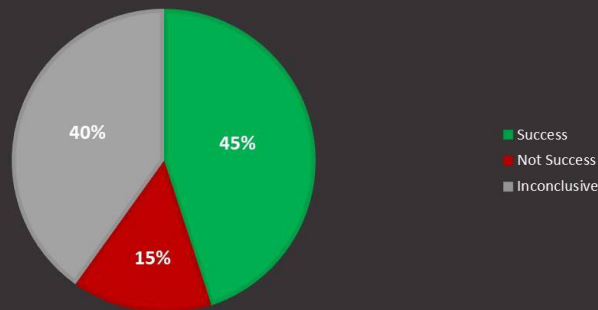
3=It is not related to your long-term career goals.

4=You do not have long-term career goals.

Algorithm to Determine Success

```
s = F3B34(job satisfaction)(1-5)
g = F3B29(job relate to goals)(1-4)
boolean success(){
.....
  match s with
  | nil -> inconclusive
  | 5,4 -> false
  | 3 -> match g with
    | nil -> inconclusive
    | 3,4 -> false
    | 1,2 -> true
  | 2,1 -> true
}
```

Output Algorithm Results:



Success Algorithm Output

Total Respondents: 16,198

Survey Identifiers of Success

Limited data pool to F1 surveys. F1 surveys represent responses from 12th grader which most closely resembles the applicant pool for universities. 9700 students, each with 781 survey questions as input. 7,575,700 elements

MACHINE LEARNING

Feature Selection

Each student has many "features" (their survey responses). To narrow our input size down from 781 features, we used

ANOVA (Analysis of Variance) to determine the relevance of each survey question to students' post-graduate success. Our top three relevant features were:

- Hours per week spent on extra-curricular activities (F1S27)
- Number of school-sponsored events participated in (F1XTRACU)
- Self-Efficacy in math (F1MATHSE)

Building a Model

Built a model to predict success based on survey answers as input and our algorithm results as the output. Tested accuracy by using different classifiers which used all the features, (LDA and MLP), or only the ANOVA-selected features.

Model	Mean Accuracy (%)
Linear Discriminant Analysis (LDA)	70.76
Multi-Layer Perceptron (MLP)	73.66
ANOVA LDA	74.89
ANOVA MLP	74.86
ANOVA K-Nearest Neighbors	74.84

RESULTS

As a result of our application of machine learning and data analysis, we were able to find a correlation between 12th grade habits and success. This can help universities identify incoming students that may need support finding success after graduation. With the given data, we were able to reach 75% accuracy on determining future success. This can be made more accurate with additional test cases, more complete data, and clearer indication of future success.

