

Predicting Success In College A Network-Based Machine Learning Approach

By: Patricia Samantha Orozco, Noelani Gonzalez-Villanueva & Dr. Sukhwa Hong

Introduction

Predicting the success of a student graduating is important because it helps direct their success. This creates a short or long term goal for students, in which is both attainable and measurable. Employing this prediction will also guide students to keep track of the courses needed to graduate, and do so within the time frame suggested.

This data also allows administrators to see where critical areas are within students academics and strengthen the areas to ensure a successful graduation. In addition, sentences in a language follows a structure that is similar to the pathways that students take to complete their degree. Like the English language, students are allowed to structure their pathway to their liking, but data suggests there are more efficient ways to doing so.

Research Objectives

One of the objectives of this project is to identify students pathways to graduation and see how they are connected in a network. For this connection, we are looking into the pathways that students took, their end results, and assembling the conclusion on their connection to each other.

- Through these connections, we will be able to identify what factors results in a successful graduation. The network will then also allow us to see where students face issues when the outcome produces to be an unsuccessful graduation.

Another objective is utilizing an algorithm, doc2vec, to represent text as numbers (vectors) to perform statistical analysis with computers or mathematical techniques. This will help us model sentences as vectors, which then predicts whether a student will graduate or not.

- Applying doc2vec and word2vec will help us to see how these vectors can be used to represent students and classes. This will result in providing us the connection between students' outcomes and the courses taken and grades earned throughout their years in college.

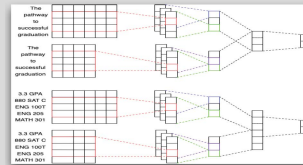
Methodology

Doc2vec is a model that represents each document as a dense vector, in which is trained to predict words in the document. By extending the idea of word2vec, doc2vec is able to learn from data which are not labeled, and take into consideration the word order in the document, such as sorting words out by frequency, or discarding rare words that appear in the document.

- By applying this model to the representation of students and taking the sequence of students' courses, this will help doc2vec generate tasks, such as finding similarity between sentences, paragraphs, documents, and data.

When constructing the prediction on whether a student will graduate or not, doc2vec will create a numeric representation of our data. This will then allow us to see visuals on the similarities between students and vectors and show how the sequence of courses students take is similar to sentences in languages.

- The spreadsheet we used for this project was a data full of certain courses students took, their GPA in the beginning of their first year, to their graduation, if graduated.



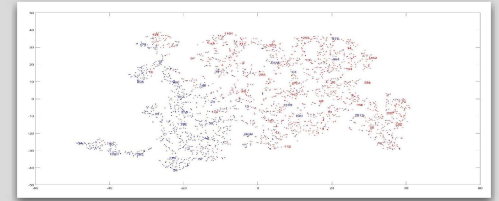
Results and Discussion

Through our research, we found that factors, such as high school GPA, SAT scores, writing intensives taken, and grades received, display a structure that is similar to that of a sentence structure. The network allows us to see a clearer path to predict graduation.

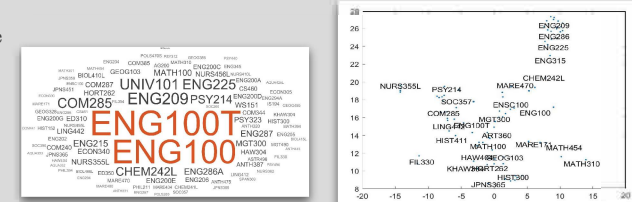
- With this structure we can predict the probability of a students projected graduation using the similarities from past student data.

Limitations

The data that was used in our research consisted of WI courses that students took throughout their academic career. With more data such as all courses taken throughout a students academic career, we will be able to improve our outcomes.



Students: graduated (blue) not graduated (red)



Student WI courses wordcloud

WI Student Course Plot

Field	Accuracy	Precision	Recall	F-score
Field 1	0.8920	0.8690	0.8618	0.8920
Field 2	0.9214	0.8740	0.8649	0.8920
Field 3	0.8881	0.8630	0.8877	0.8876
Field 4	0.8720	0.7853	0.8467	0.8168
Field 5	0.9125	0.8418	0.8619	0.8826
Field 6	0.9872	0.8888	0.8833	0.9812
Field 7	0.9541	0.8804	0.8952	0.9120
Field 8	0.8858	0.8223	0.8228	0.8228
Field 9	0.9680	0.7081	0.8603	0.8240
Field 10	0.9503	0.8187	0.8777	0.8800
Average	0.9265	0.8398	0.8703	0.8824

Precision and Recall: 10% test 90% training test

Conclusion

By using node2vec, we can predict the likelihood of graduation by representing each student as a node. The network of nodes displays how students achieved graduation with their path. If a student shows a similar pathway to that of a student who graduated, we can predict the probability of the outcome to be similar to that of past students. To predict the probability of graduation, we consider a student's high school GPA, SAT scores, writing intensives taken, and grades received.