

Thinkful - Final Capstone Proposal

Predicting worthiness of college education using machine learning

Introduction:

There has been a discussion about value of college education in the US due to the rising student loan problem as well as offering of alternative education services through increase in online education system. Considering this, there is a value in evaluating if attending college will be worth it for an individual by looking at earnings to student loan ratio of graduates attending US colleges.

The problem:

Rising student loan debt has casted a doubt regards to value of a college education. This tool will allow people to estimate worthiness of attending a college by looking at the earnings to student loans ratio.

Value:

The tool is valuable as it would provide a way to measure how worthy college investment would be by looking at the cost / benefit ratio of attending college. It would also provide a way to estimate earnings based on what college you attend as a measure to evaluate if attending that college would be worth the cost.

Data Source:

For this project, I will access by data through College Scorecard which has data on college attendees from 1996 - 2017 with information such as student completion, debt and repayment, earnings, and other data points. (Source: <https://collegescorecard.ed.gov/data/>)

Tools/ Techniques:

Techniques that I plan on using are unsupervised learning as well as Keras and TensorFlow. I also plan on seeing how I can implement supervised learning to answer this question.

Challenges:

The biggest challenge I see myself facing is the lack of pre-labeled data. I am not entirely sure how I can use this dataset in regards to supervised learning as each dataset is not already pre-labelled as worth the investment vs not. I will also have to define what is worth the investment. I plan on doing this by creating a percentage of debt to earnings ratio and labelling those with certain percentage as "worth it" versus others as "not worth it".

Limitations:

There are some clear limitations such as social value of attending college. Undergraduate degrees are prerequisites to many office jobs. It also allows people to have a network hence, it is difficult to measure value just based on earnings / debt ratio.