

Big data Project Proposal

March 11, 2019

Idea

It is a current kaggle competition for recommending questions to appropriate volunteers(professionals) ,so our main role is to develop a method to recommend relevant questions to the professionals who are most likely to answer them .

Problem statement

The U.S. has almost 500 students for every guidance counselor. Underserved youth lack the network to find their career role models, making CareerVillage.org the only option for millions of young people in America and around the globe with nowhere else to turn.

To date, 25,000 volunteers have created profiles and opted in to receive emails when a career question is a good fit for them. This is where your skills come in. To help students get the advice they need, the team at CareerVillage.org needs to be able to send the right questions to the right volunteers. The notifications sent to volunteers seem to have the greatest impact on how many questions are answered.

Dataset(s)

1. **Kaggle competition datasets :**

<https://www.kaggle.com/c/data-science-for-good-careervillage/data>

Proposed solution

We will build Hybrid engine recommender that combines content based filtering and Collaborative Filtering

1. Content based filtering

In this module the recommendations are based on the similarities between the question asked by a student and the pre-answered questions by the professionals in the system .

2. Collaborative filtering

In this module the recommendations are based on the idea that professionals similar to another professional can be used to predict how much this professional will answer a particular question even if he didn't answer a similar question before .

3. Hybrid engine

We will combine the two previous approaches to build a hybrid engine recommender .

4. Restricted Boltzmann machine (Optional)

We will try to use Restricted Boltzmann machine if the accuracy isn't high .