

# **PREDICTION OF STUDENT EARNINGS AND DEBTS**

Kruti Chauhan

## **Problem:**

Many students have the notion that graduation leads to better financial status. However, the changing economy has led to more education loan debts questioning the basis fact stated above. Hence, machine learning techniques can be applied to predict the earnings and debts of the alumni based on numerous factors such as entrance tests, acceptance rate, ranking et cetera. This will indeed help students make an informed decision about the selection of a college.

## **Data Sets:**

The US Department of Education has released a data set [1] which includes information of about 8000 colleges with around 2000 data fields.

## **Methodology:**

The data set contains a considerable amount of missing values. Dimensionality reduction method such as PCA applied to reduce the feature space.

Regression algorithms such as neural networks, SVR and Lasso to predict the earnings and debt of a student of a particular college.

## **Experiments:**

This methodology can be validated using cross-validation as 60-20-20 for train-validation-test. MSE would be used to evaluate the methodology and pick the method obtaining the least error MSE.

## **Related work and Novelty:**

Prediction of earnings and debts for a college was done to find correlations between characteristics of colleges and the future success of their alumni has great value to society[2]. This paper used regression method to fill the missing values, and hence regressors such as SVR, KNN and Neural networks were applied. Another paper[3] which predicted the earnings, used mean imputation method to fill the missing values and later on applied regressors such as Ridge, Lasso, and Random Forest.

Handling missing values is a challenging task for this data set and can be handled by numerous methods [4]. This project hence additionally, would apply and experiment these methods.

## References:

1. U.S. Department of Education. College Scorecard, 2015, <https://collegescorecard.ed.gov/data/>
2. Agrawal, Monica, Priya Ganesan, and Keith Wyngarden. "Prediction of Post-Collegiate Earnings and Debt."
3. Strand, Miranda, and Tommy Truong. "Predicting Student Earnings After College."
4. Marlin, Benjamin M. *Missing data problems in machine learning*. Diss. University of Toronto, 2008.
5. Rubin, Donald B. "Multiple imputation after 18+ years." *Journal of the American statistical Association* 91.434 (1996): 473-489.