CS 6375.001: Machine Learning
Project Proposal: Predicting an employee's wage

Submitted By:

Falak Singhal (fxs161530)
Melvin James (mxj162130)
Vidya Sri Mani (vxm163230)

INTRODUCTION

The annual wage of an employee is influenced by several factors like occupation, position, location etc. Often, a college education results in a higher pay than for those who do not have a college degree. Several factors pertaining to the university such as acceptance rate, average score also factor in while determining the wage. This project proposes to accurately predict the pay of an employee by considering various factors that might impact the result.

DATASET AND FEATURES

For this project, we intend to use the Adult dataset which is sourced from the University of California Irvine data repository. The data set is available under public licenses and open source to be used by the machine learning community for the empirical analysis of machine learning algorithms. It contains information about an individual's age, level of education and current employment type etc. This multivariate dataset has close forty-nine thousand records and a binomial label which indicates whether the salary of the employee is lesser than or greater than fifty thousand dollars. The dataset can be found at the URL –

http://archive.ics.uci.edu/ml/datasets/Adult

A total of fourteen attributes are mentioned in this dataset, seven of which are polynomials, one binomial and six continuous attributes. Seventy six percent of the records in the dataset have a class label of <50K. The training set consists of thirty-two thousand records while test dataset consists of sixteen thousand records.

HELPFUL PAPERS

- Ron Kohavi, "Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid", Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996
- Ron Kohavi and Barry G. Becker and Dan Sommerfield. "Improving Simple Bayes". Data Mining and Visualization Group Silicon Graphics, Inc
- Rumberger, R. W., Thomas S. L. The economic returns to college major, quality and performance: A multilevel analysis of recent graduates. Economics of Education Review, 12(1):1–19, 1993.

TIME-FRAME

Anticipated duration of project: March 15, 2017 – April 30, 2017

TEAMMATES

- Falak Singhal (fxs161530)
- Melvin James (mxj162130)
- Vidya Sri Mani (vxm163230)