

University of Colorado at Colorado Springs
Home Work Assignment 1: Data Mining and Feature Engineering
Out: 01-23-2019, Due: every day

1 Preface

In this homework assignment, you will learn how to manipulate data for Machine Learning. Using the Colorado Voter Registration dataset you will explore feature engineering and data visualization.

2 Assignment

Write a program to download the Colorado Voter Registration dataset. Write a program to prepare the dataset for analysis by Machine Learning algorithms. Explain all that you can about the dataset. Plot the data to illustrate your conclusions and provide all the necessary statistics to support your feature engineering decisions.

Don't worry if you don't understand how to proceed; we will discuss this in class.

All of your work will be done in Python. If you need to download the dataset you can use other means (such as a shell script) but you must supply an automated way of retrieving datasets for all work. As part of this assignment you need to get python up-and-running, and demonstrate that you can use it to manipulate datasets. Use any libraries that you like for this task but all datasets for this class should first be loaded into a Pandas DataFrame. We will explain this in class.

3 Conclusions

As mentioned in the onset, please do what you have been tasked with above. Perform additional research, do additional implementations, perform additional experiments, etc., if you want Extra Credit. Be enthused about Regression and Machine Learning!