

CS 210 Project Blog

MENU

Reports

To make it easier, I have written a report consisting of Introduction, Part1 and Part2.

Below, you can find the pdf's for these reports:

[Project Overview \(Abstract\)](#)

[Part 1 Overview](#)

[Part 2 Overview](#)

May 19, 2017 / [Leave a comment](#)

ZeroR Update

I've calculated ZeroR's again.

You can find the code in GitHub.

May 19, 2017 / [Leave a comment](#)

Equal Weight DTCs

I've constructed DTC's with equal weights. For that I used these 2 approaches:

1. I've randomly selected the "No" labels as equal quantitatively as "Yes".
2. I've applied k-means to "No" labels with $k = \#$ of "Yes" labeled, and then took the centroids of all k clusters.

Then I used the same DTC classifier and constructed DTCs.

You can find the code on GitHub.

May 14, 2017 / [Leave a comment](#)

Plan for last phase

Up to now, I have completed the following steps in part 2:

- Linear Regression to predict the exact grade
- Decision Tree, SVM, Logistic Regression to predict the grade in 5 groups
- Decision Tree, Logistic Regression to predict Top/Bottom 10%–20%

Now, I'll do the following step:

- The classification for top or bottom 10–20% give bad results since there are lots of no's and a few yes's. With standard weights, there are lots of false negatives. Therefore, my next step is to modify weights so that they have equal weight. By doing so, I'll smooth the false positive and false negative numbers.

May 12, 2017 / [Leave a comment](#)

Logistic Regressions for Percentages

I've applied Logistic Regressions to predict:

- Top 10% or Not
- Top 20% or Not
- Bottom 10% or Not
- Bottom 20% or Not

Code: [MatPercentLogisticRegression.pdf](#) [PorPercentLogisticRegression.pdf](#)

May 7, 2017 / [Leave a comment](#)

Decision Trees for Percentages

I've constructed decision trees to predict:

- Top 10% or Not
- Top 20% or Not
- Bottom 10% or Not
- Bottom 20% or Not

Code: [MatPercentDecisionTree.pdf](#) [PorPercentDecisionTree.pdf](#)

May 7, 2017 / [Leave a comment](#)

Processing Data for Percentages

I've sorted the data and identified Top 10%, Top 20%, Bottom 10% and Bottom 20%.

In the next step, I will try to predict these without any test scores.

Code: [Percentages.pdf](#)

May 7, 2017 / [Leave a comment](#)



RECENT POSTS

- [Reports](#)
May 19, 2017
- [ZeroR Update](#)
May 19, 2017
- [Equal Weight DTCs](#)
May 14, 2017
- [Plan for last phase](#)
May 12, 2017
- [Logistic Regressions for Percentages](#)
May 7, 2017
- [Decision Trees for Percentages](#)
May 7, 2017
- [Processing Data for Percentages](#)
May 7, 2017
- [Confusion Matrices](#)
May 3, 2017
- [Comparison of Different Models](#)
April 29, 2017
- [Logistic Regressions](#)
April 29, 2017
- [SVMs](#)
April 29, 2017
- [Decision Trees](#)
April 29, 2017
- [Linear Regressions](#)
April 29, 2017
- [Getting and Processing Data](#)
April 29, 2017
- [GitHub Integration](#)
April 26, 2017
- [Part 1 Overview](#)
April 24, 2017
- [Planning for the rest of the project](#)
April 24, 2017

- [Regression \(New\)](#)
April 14, 2017
- [Cluster](#)
April 13, 2017
- [Regressions and Some Plots](#)
April 9, 2017
- [Minimized Data](#)
April 8, 2017
- [Correlation Matrix & Scatter Plot Matrix](#)
March 24, 2017
- [Filtered Missing Data](#)
March 24, 2017
- [Merged Data](#)
March 23, 2017
- [Code for Histograms](#)
March 18, 2017
- [Histograms](#)
March 18, 2017
- [Data Downloaded](#)
March 18, 2017
- [Project Proposal \(v3\)](#)
March 18, 2017
- [Project Proposal \(v2\)](#)
March 2, 2017
- [Project Proposal](#)
March 1, 2017