



Internship Report

2016 summer

Shu Zhang

@ Huawei Technology

My Internship Project

Optimize the application performance based on machine learning models
My main work : analysis of offline data. (AWS S3)

DATA

7 input features

3 output target values

Y := function(3 outputs)

*. CSV

429	0.33333	0.142857	0.33333	0.57142857	0	34.168	0	20.5
1	0.33333	0	0.33333	0.71428571	0	51.952	0	21
1	0.66667	0.285714	0	0.57142857	0.33333333	65.493	0	36
857	0.66667	0.142857	0.16667	0.57142857	0.66666667	70.367	0	29.6
714	0.16667	0	0	0.42857143	0.33333333	47.349	0	25.8
857	0	0	0	0.28571429	0.66666667	53.984	0	150
571	0	0.142857	0	0.14285714	0.33333333	36.109	100	####
714	0	0.285714	0	0	0.66666667	40.713	100	####
571	0	0	0	0.14285714	1	43.195	100	####
429	0	0	0.16667	0	0	23.018	100	####
571	0.33333	0	0	0.28571429	0.33333333	38.501	0	61.8

Machine Learning Models

Regression

Y are continuous variables.

- Support vector machine
- Root mean square error to check the results

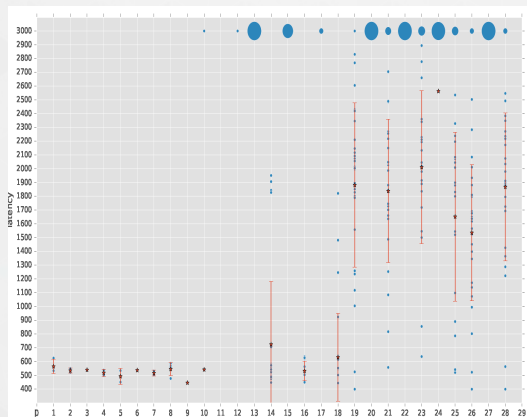
Classification

Label Y as 0,1

- Decision Tree, SVM, Logistic Regression
- Confusion matrix to check the accuracy

Visualization

- Histogram
- Bubble Scatter Plot
- Error Bar
- Line chart



Program that I used

Python

<https://www.codecademy.com/learn>

Pandas, Numpy, scikit-learn, matplotlib

codecademy

R

ggplot2



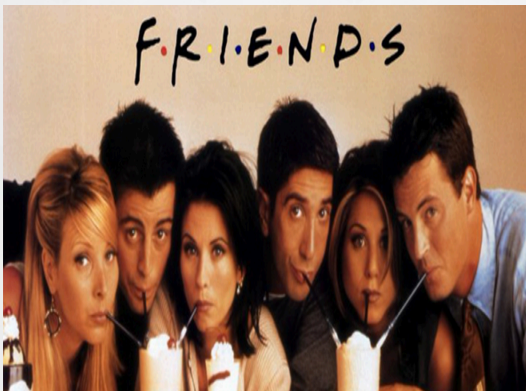
Git

Is a free and open source distribution version control system

Github



Finding the internship



Thank you