feature:

1. submission\_list( former submissions of a image)

2. title\_len\_list (the title length)

3. raw\_time\_list ( four time periods during a day, and the timezone(-7,-8, a vector containing 6 binary elements)

dataset:

1. train set : test set = 3 : 1 (99230 :33077)

2. 10 fold cross validation on train set

3. Normalize: StandardScalar

predict label :comment number of a post

A. binary classification:

**a) whether or not the submission is good**

1 : data[i]['number\_of\_comments']\*0.2 + data[i]['total\_votes']\*0.8 > 36 (large)

0. otherwise (not large)

**b) model used**

# 1. KNN （K =5）

**validation**

Accuracy: 0.60 (+/- 0.01)

Recall: 0.60 (+/- 0.01)

Precision: 0.60 (+/- 0.01)

**test**

precision: 60.16%,

recall: 59.57%

accuracy: 60.33%

# 2. NB

**validation**

Accuracy: 0.63 (+/- 0.01)

Recall: 0.63 (+/- 0.01)

Precision: 0.64 (+/- 0.01)

**test**

precision: 60.19%,

recall: 76.79%

accuracy: 63.24%

# 3. LR

**validation**

Accuracy: 0.64 (+/- 0.01)

Recall: 0.64 (+/- 0.01)

Precision: 0.65 (+/- 0.01)

**test**

precision: 61.47%,

recall: 74.14%

accuracy: 64.07%

Theta :

[-0.40427441, -0.03100889, 0.03594161, 0.00226282, 0.02870109, 0.01993202, 0.05955811, 0.50435371]

# 4. RF

**validation**

Accuracy: 0.63 (+/- 0.01)

Recall: 0.63 (+/- 0.01)

Precision: 0.63 (+/- 0.01)

**test**

precision: 61.54%,

recall: 69.67%

accuracy: 63.31%

# 5. DT

**validation**

Accuracy: 0.66 (+/- 0.01)

Recall: 0.65 (+/- 0.01)

Precision: 0.66 (+/- 0.01)

**test**

precision: 63.48%,

recall: 72.45%

accuracy: 65.61%

# 6. ADA

**validation**

Accuracy: 0.66 (+/- 0.01)

Recall: 0.65 (+/- 0.01)

Precision: 0.66 (+/- 0.01)

**test**

precision: 63.90%, recall: 70.51%

accuracy: 65.56%