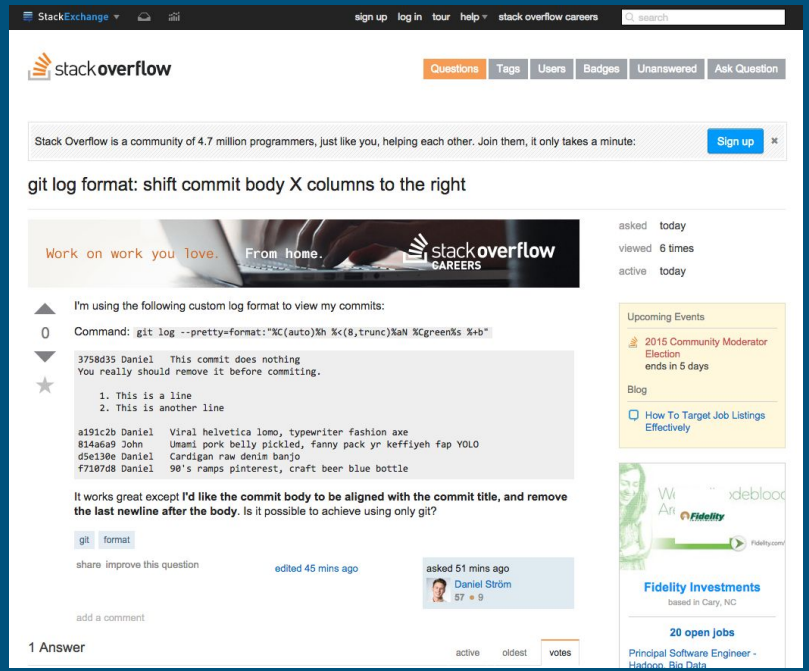
SE text mining in StackExchange

Zhe

Stackoverflow



Tasks

a. Generate tags for posts on Stackoverflow [1]
b. Find most frequent topics on Stackoverflow (LDA, unsupervised) [2]
c. Retrieve existing discussions on Stackoverflow given context [3]

[1] Stanley, C., & Byrne, M. D. (2013). Predicting tags for stackoverflow posts. *In Proceedings of ICCM* (Vol. 2013).
[2] Wang, S., Lo, D., & Jiang, L. (2013, March). An empirical study on developer interactions in StackOverflow. *In Proceedings of the 28th Annual ACM Symposium on Applied Computing* (pp. 1019-1024). ACM.
[3] Ponzanelli, L., Bavota, G., Di Penta, M., Oliveto, R., & Lanza, M. (2014, May). Mining stackoverflow to turn the IDE into a self-confident programming prompter. *In Proceedings of the 11th Working Conference on Mining Software Repositories* (pp. 102-111). ACM.

Generate tags for posts on Stackoverflow

a. Clayton and Byrne, 2013 [1]. ACT-R inspired Bayesian probabilistic model. 65% accuracy.
b. Kuo, 2011 [2]. Co-occurrence model. 47% classification accuracy.
c. Fu & Pirolli, 2007 [3]. SNIF-ACT. No result on Stackoverflow.

[1] Stanley, C., & Byrne, M. D. (2013). Predicting tags for stackoverflow posts. *In Proceedings of ICCM* (Vol. 2013).
[2] Kuo, D. (2011). On word prediction methods. *Technical report*, EECS Department, University of California, Berkeley.
[3] Fu, W. T., & Pirolli, P. (2007). SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Human–Computer Interaction*, 22(4), 355-412.

Why accuracy

Evaluation methods of multi-class tasks [1]

Measure	Formula	Evaluation focus
Average Accuracy	$\frac{\sum_{i=1}^{l} \frac{tp_i+tn_i}{tp_i+fn_i+fp_i+tn_i}}{l}$	The average per-class effectiveness of a classifier
Error Rate	$\frac{\sum_{i=1}^{l} \frac{fp_i+fn_i}{tp_i+fn_i+fp_i+tn_i}}{l}$	The average per-class classification error
$Precision_\mu$	$\frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l}(tp_i+fp_i)}$	Agreement of the data class labels with those of a classifiers if calculated from sums of per-text decisions
$Recall_\mu$	$\frac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l}(tp_i+fn_i)}$	Effectiveness of a classifier to identify class labels if calculated from sums of per-text decisions
$Fscore_\mu$	$\frac{(\beta^2+1)Precision_\mu Recall_\mu}{\beta^2 Precision_\mu+Recall_\mu}$	Relations between data's positive labels and those given by a classifier based on sums of per-text decisions
$Precision_M$	$\frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i+fp_i}}{l}$	An average per-class agreement of the data class labels with those of a classifiers
$Recall_M$	$\frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i+fn_i}}{l}$	An average per-class effectiveness of a classifier to identify class labels
$Fscore_M$	$\frac{(\beta^2+1)Precision_M Recall_M}{\beta^2 Precision_M+Recall_M}$	Relations between data's positive labels and those given by a classifier based on a per-class average

[1] Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427-437.

Evaluation

a. Fscore_mu

weighted mean of F-scores of each class

Sum (F(i) * Population(i)) / Sum (Population(i))

b. Fscore_M

unweighted mean of F-scores of each class

Sum (F(i))/ Number of Classes

What can SMOTE do

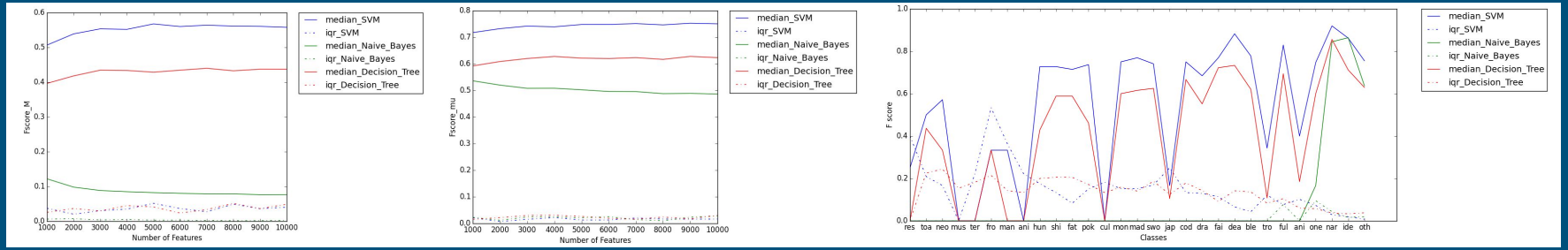
After SMOTE:

a. Fscore_mu goes down (not really maybe...)

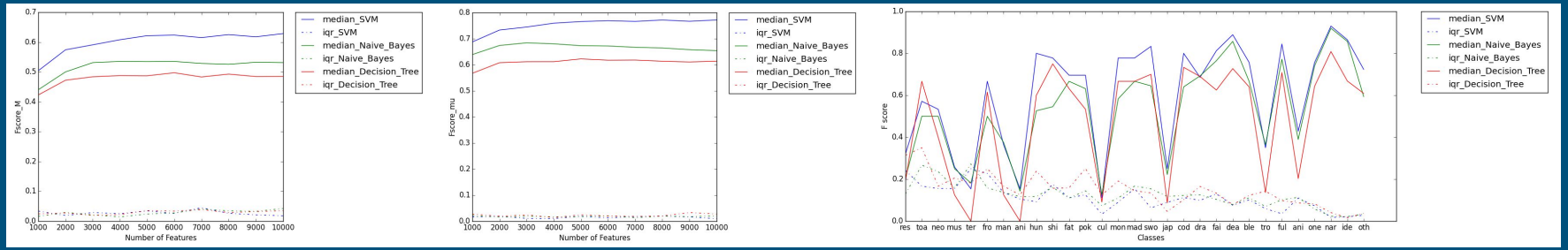
b. Fscore_M goes up (significantly)

SMOTE

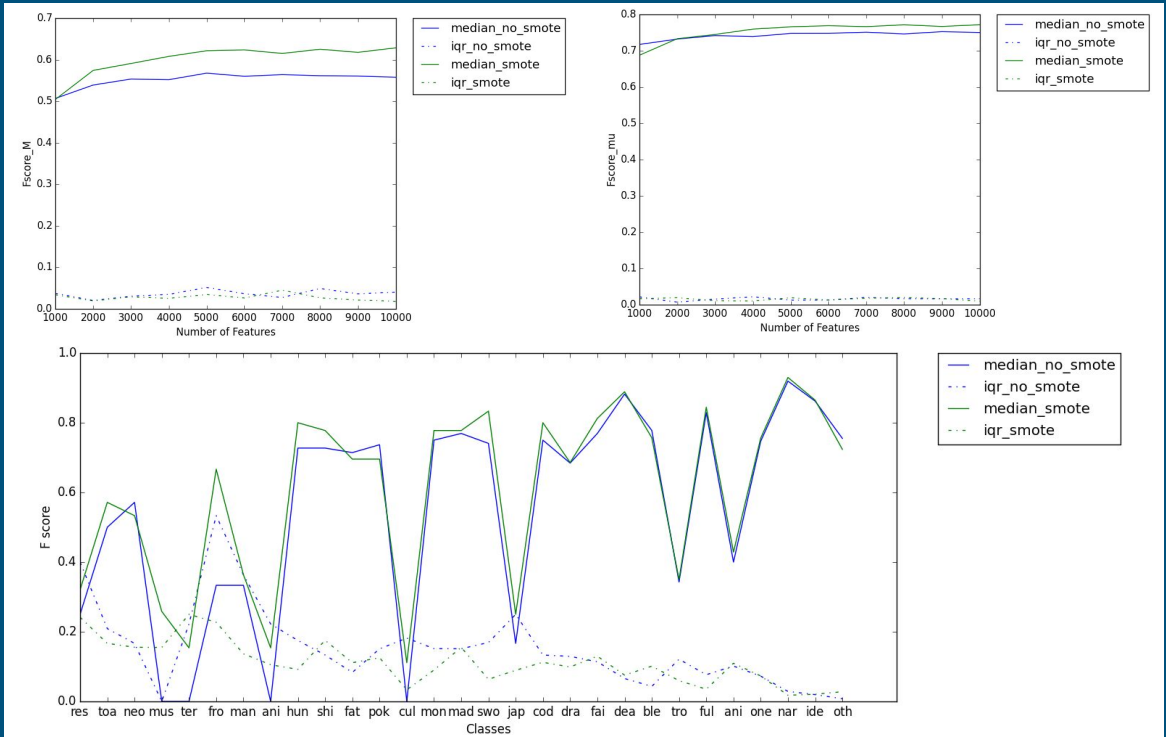
Before:



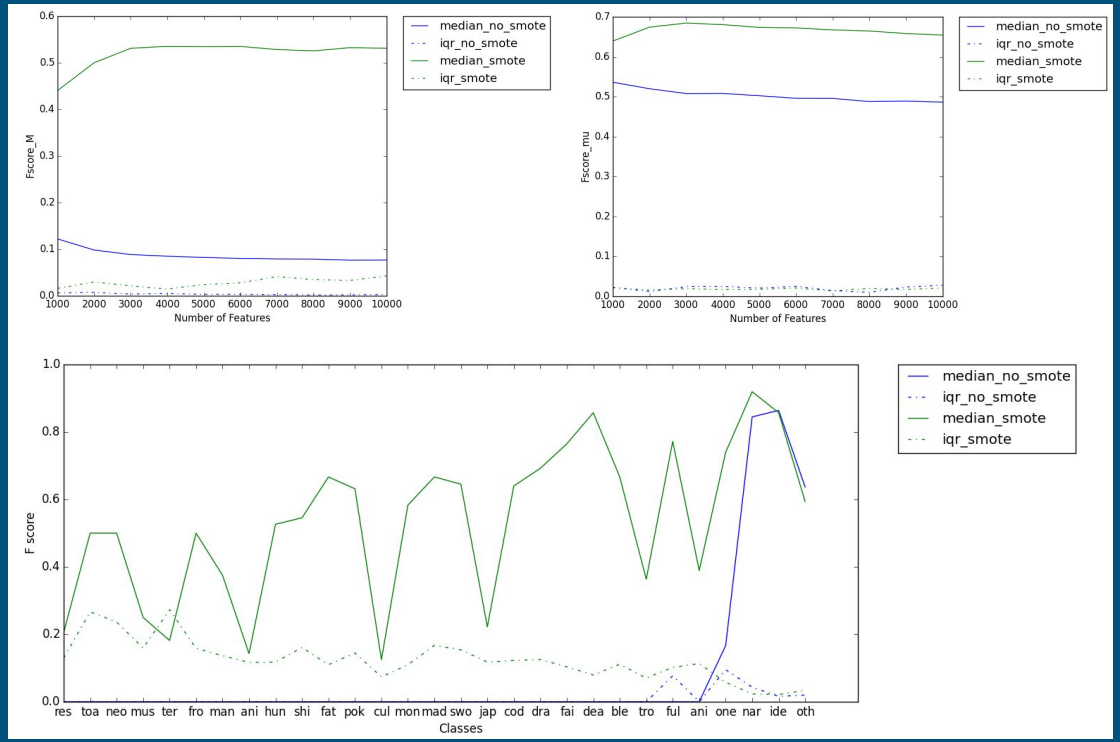
After:



SVM (Linear)


median_no_smote
iqr_no_smote
median_smote
iqr_smote
Fscore_M
Number of Features
Fscore_mu
Number of Features
F score
Classes
res toa neo mus ter fro man ani hun shi fat pok cul mon mad swo jap cod dra fai dea ble tro ful ani one nar ide oth

Naive Bayes (Multinomial)



Decision Tree (CART)