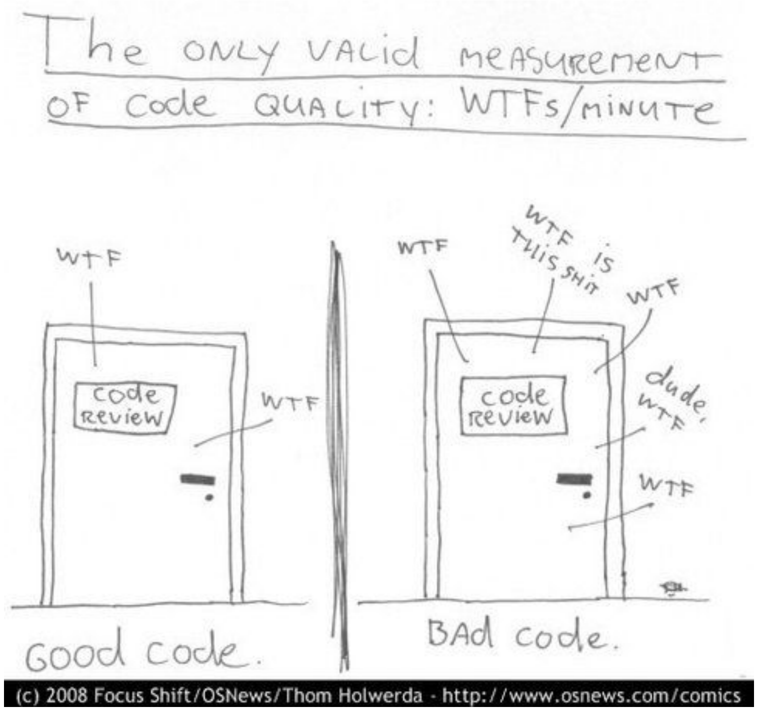
Mining Software Repositories Lab

Week2
Group 10

Motivation

- Faster Delivery
- Higher Quality
- Better Understanding of projects



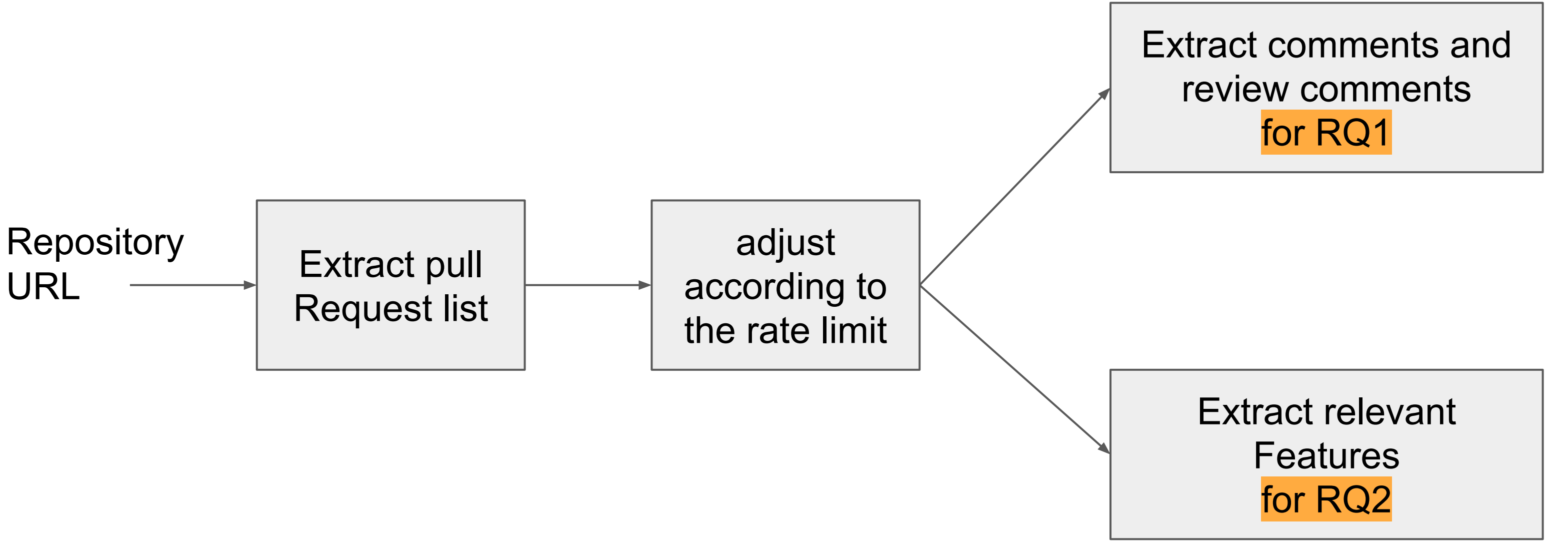
(c) 2008 Focus Shift/OSNews/Thom Holwerda - http://www.osnews.com/comics

Research Questions:

RQ1:What are the reasons for changes in pull requests?

RQ2: Can we predict whether a pull request will be merged?

Data Collection : workflow



Repository URL

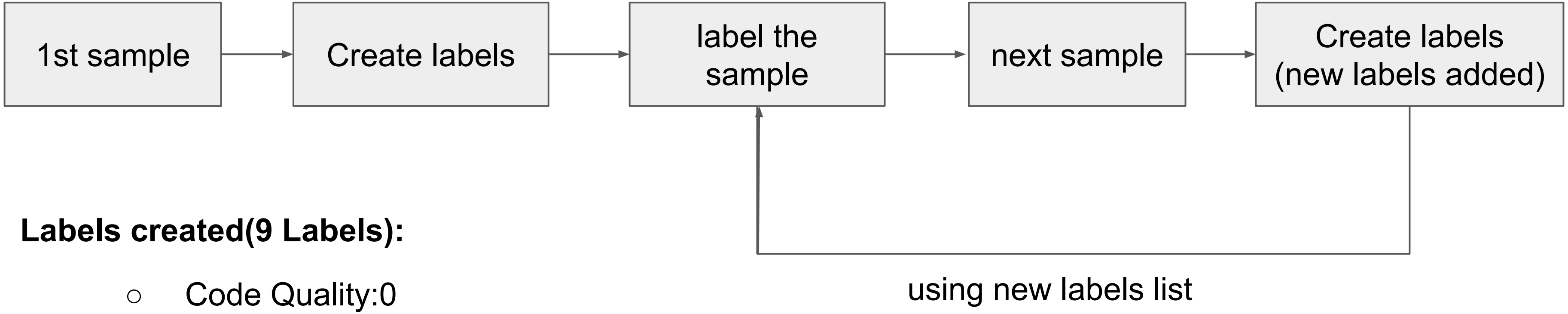
Extract pull Request list

adjust according to the rate limit

Extract comments and review comments for RQ1

Extract relevant Features for RQ2

RQ1: Inductive Coding



1st sample

Create labels

label the sample

next sample

Create labels (new labels added)

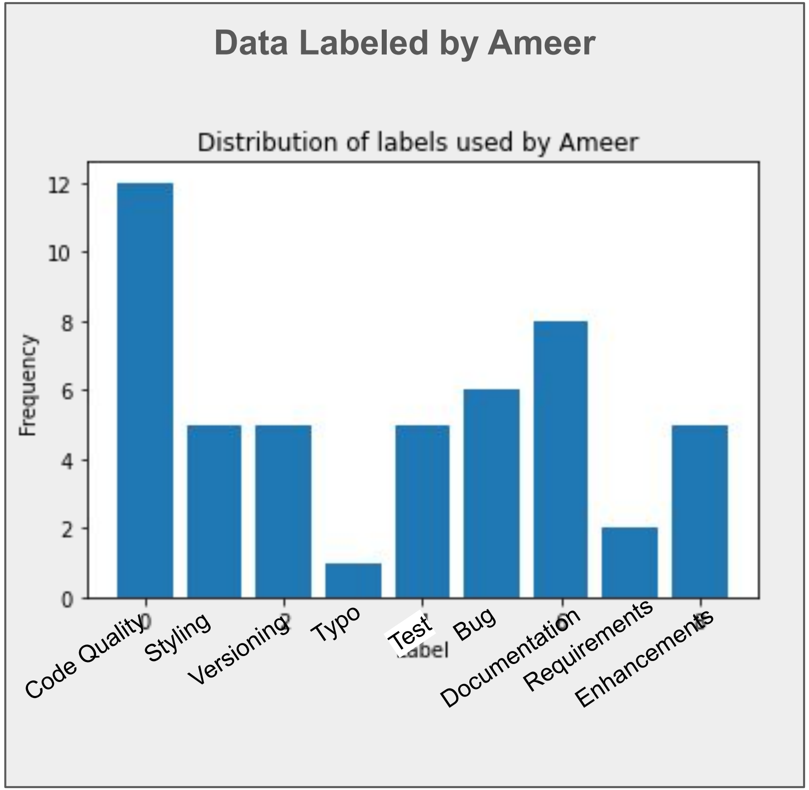
using new labels list

Labels created(9 Labels):

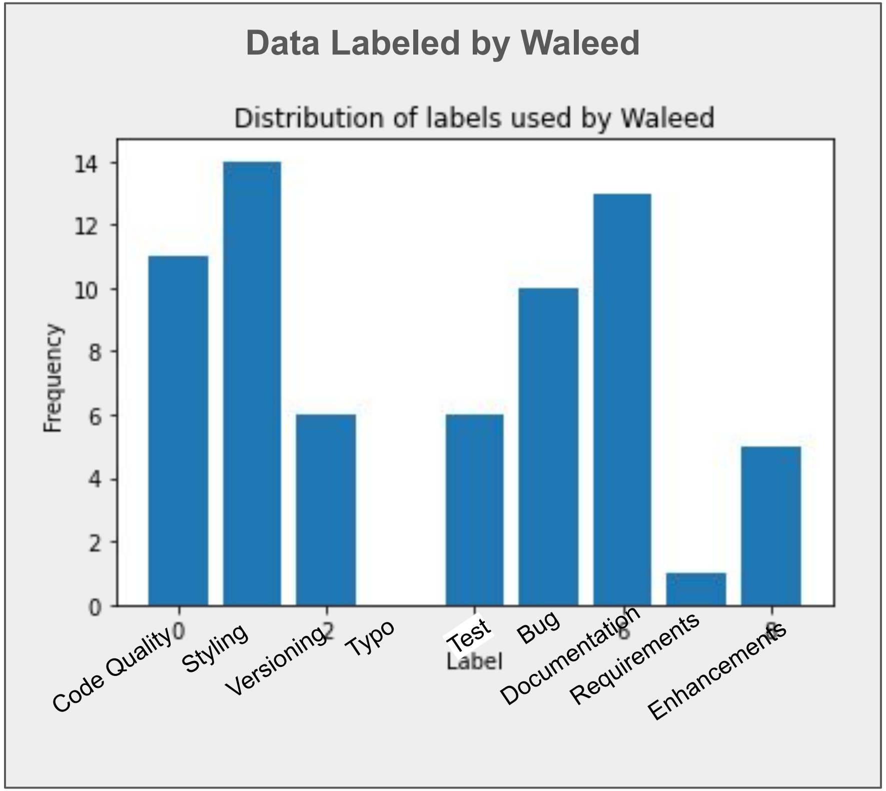
- Code Quality:0
- Styling:1
- Versioning:2
- Typo: 3
- Test: 4
- Bug:5
- Documentation:6
- Requirements: 7
- Enhancement:8

- **Sample size :** 49 pull requests
- **Repository source :** commons-configuration , commons-lang repositories

RQ1: Deductive Coding



Data Labeled by Ameer



Data Labeled by Waleed

RQ1 Qualitative analysis : Inter-rater agreement

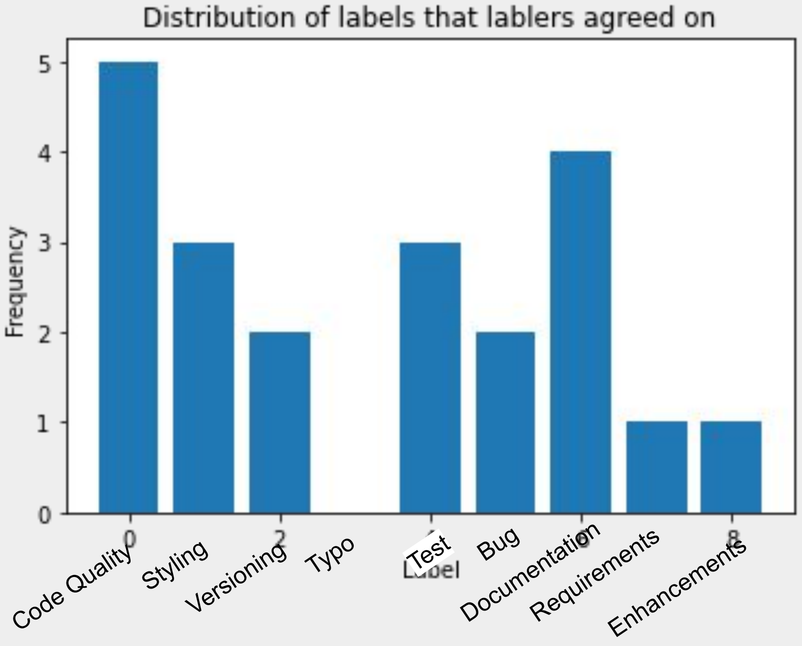
How to measure the agreement between the two annotators ?

Cohen Kappa Score

0.36

Indicates fair agreement

Agreement Sample



RQ1:What are the reasons for changes in pull requests?

Reasons / Labels:	%
Code Quality	24%
Styling	10%
Versioning	10%
Typo	2%
Test	10%
Bug	12%
Documentation	16%
Requirements	4%
Enhancements	10%

RQ2: Can we predict whether a pull request will be merged?

Size: 3321

Distribution :

- merged (True) : 2015
- non-merged (False): 1306

Features created :

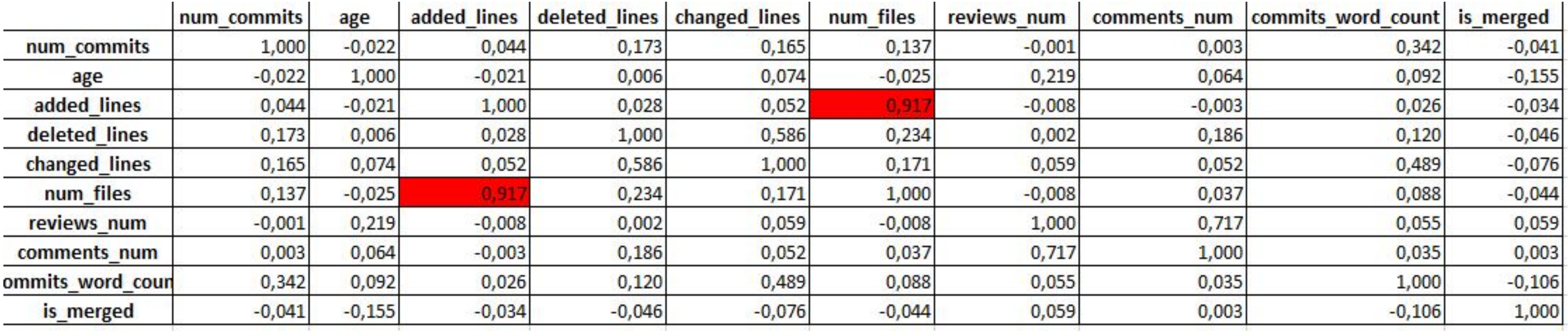
- **num_commits** : Number of commits per Pull Request.
- **Age** : number of days since creation of the repository
- **added_lines**: Lines added in Pull request, from all files.
- **deleted_lines**: Lines added in Pull request, from all files.
- **changed_lines**: Lines changed in Pull request, from all files.
- **num_files** : Number of lines edited.
- **reviews_num**: number of reviews on pull request.
- **comments_num**: Number of comments on the Pull request.
- **commits_word_count :** Number of words in the commit

Data Collection : Results for RQ2

Repository	Number of PR	Language	Starts	Forks	Contributors
alibaba/arthas	700	Java,	31.9k	6.9	169
vuejs/router	955	JavaScript	2.7k	896	179
apache/commons-math	222	Java	492	339	52
nodejs/node-gyp	679	JavaScript	8.9k	1.7k	211
adamchainz/django-mysql	765	Python	516	106	28

RQ2: Feature correlation

Correlation Matrix



	num_commits	age	added_lines	deleted_lines	changed_lines	num_files	reviews_num	comments_num	commits_word_count	is_merged
num_commits	1,000	-0,022	0,044	0,173	0,165	0,137	-0,001	0,003	0,342	-0,041
age	-0,022	1,000	-0,021	0,006	0,074	-0,025	0,219	0,064	0,092	-0,155
added_lines	0,044	-0,021	1,000	0,028	0,052	0,917	-0,008	-0,003	0,026	-0,034
deleted_lines	0,173	0,006	0,028	1,000	0,586	0,234	0,002	0,186	0,120	-0,046
changed_lines	0,165	0,074	0,052	0,586	1,000	0,171	0,059	0,052	0,489	-0,076
num_files	0,137	-0,025	0,917	0,234	0,171	1,000	-0,008	0,037	0,088	-0,044
reviews_num	-0,001	0,219	-0,008	0,002	0,059	-0,008	1,000	0,717	0,055	0,059
comments_num	0,003	0,064	-0,003	0,186	0,052	0,037	0,717	1,000	0,035	0,003
ommits_word_coun	0,342	0,092	0,026	0,120	0,489	0,088	0,055	0,035	1,000	-0,106
is_merged	-0,041	-0,155	-0,034	-0,046	-0,076	-0,044	0,059	0,003	-0,106	1,000

num_files and added_lines are highly correlated
so we can remove one of them

we removed num_files

RQ2: Modeling

10-Fold Cross validation

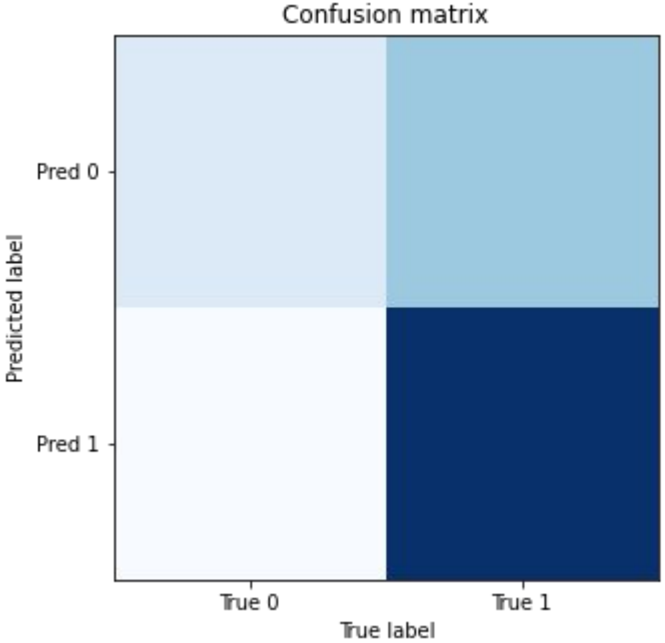
Model Name	Mean accuracy
Decision Tree Classifier	0.62
K Neighbors Classifier	0.63
Logistic Regression	0.62
Gaussian Naive Bayes	0.61
Bernoulli Naive Baye	0.6

Chosen model : K Neighbors Classifier

RQ2: Model evaluation

Data splitting :
- 80% Train
- 20% Test

Precision	0.64
Recall	0.79
Accuracy	0.63
F1-score	0.71



RQ2: Feature Engineering

no ‘Age’

Precision	0.62
Recall	0.71
Accuracy	0.58
F1-score	0.66

‘Age’

Precision	0.64
Recall	0.79
Accuracy	0.63
F1-score	0.71

commits_word_count’ had no effect on the model

RQ2: comparing the model to a baseline

Baseline random classifier with a distribution : 40% non_merged , 60% merged

Performance evaluation
of the baseline

Precision	0.60
Recall	0.6
Accuracy	0.51
F1-score	0.60

Performance evaluation
of our model

Precision	0.64
Recall	0.79
Accuracy	0.63
F1-score	0.71

RQ2: Can we predict whether a pull request will be merged?

Yes !

Conclusion and Discussion

- We can find more reasons if we study more pull requests from different repositories
- Labeling the pull requests with the reasons can be automated using NLP and machine learning
- We may get better agreement score if we consider having at least a common label in the labels set for the two labelers
- We can improve the model by extracting more data and add the PR labels as a feature.

References

https://arxiv.org/pdf/2105.13970.pdf

https://sophilabs.com/blog/pr-prediction-machine-learning

Thank you for your attention