

Mining Software Repositories Lab

Group 10

Week 1

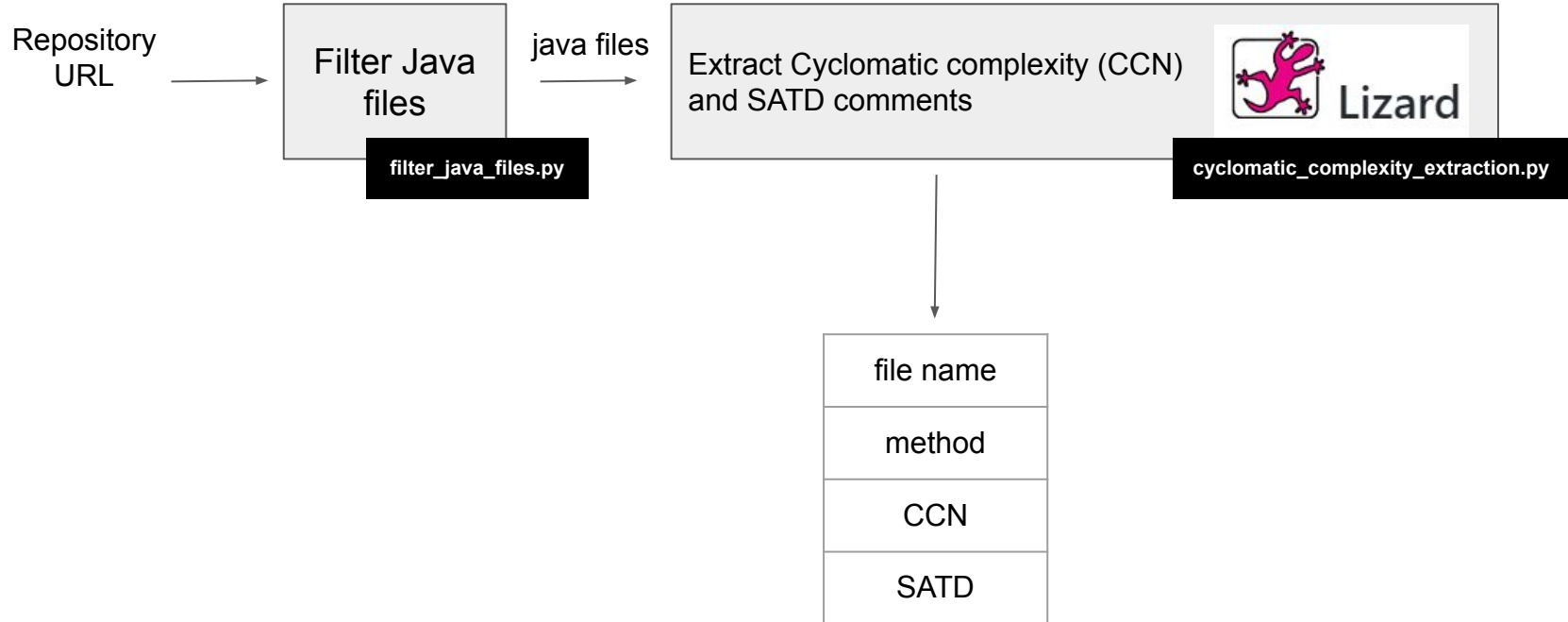
Motivation

- Static file analysis
- Better Developer tools higher efficiency
- Find Issues as early as possible

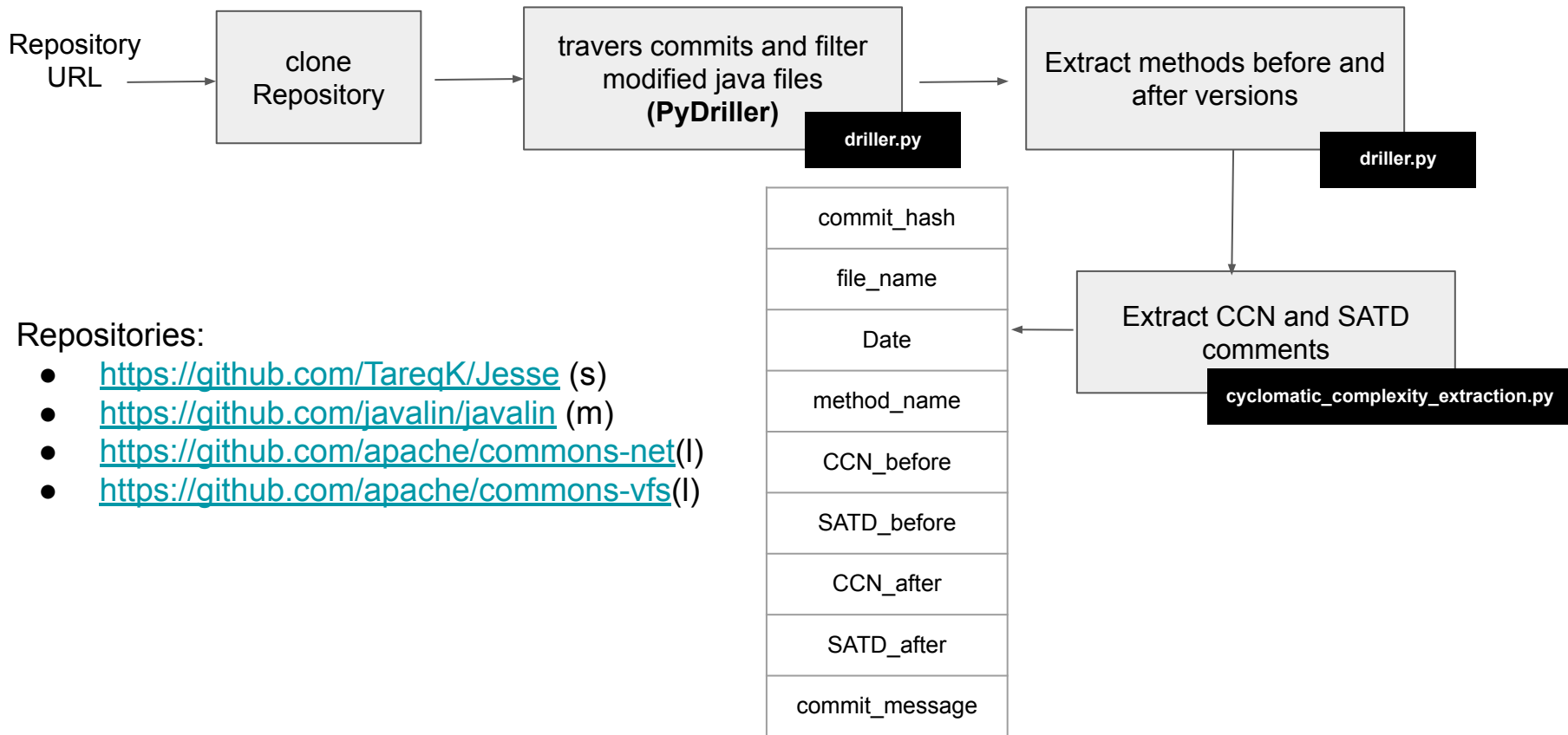
RQ1 Does SATD in code comments correlate with high code complexity?

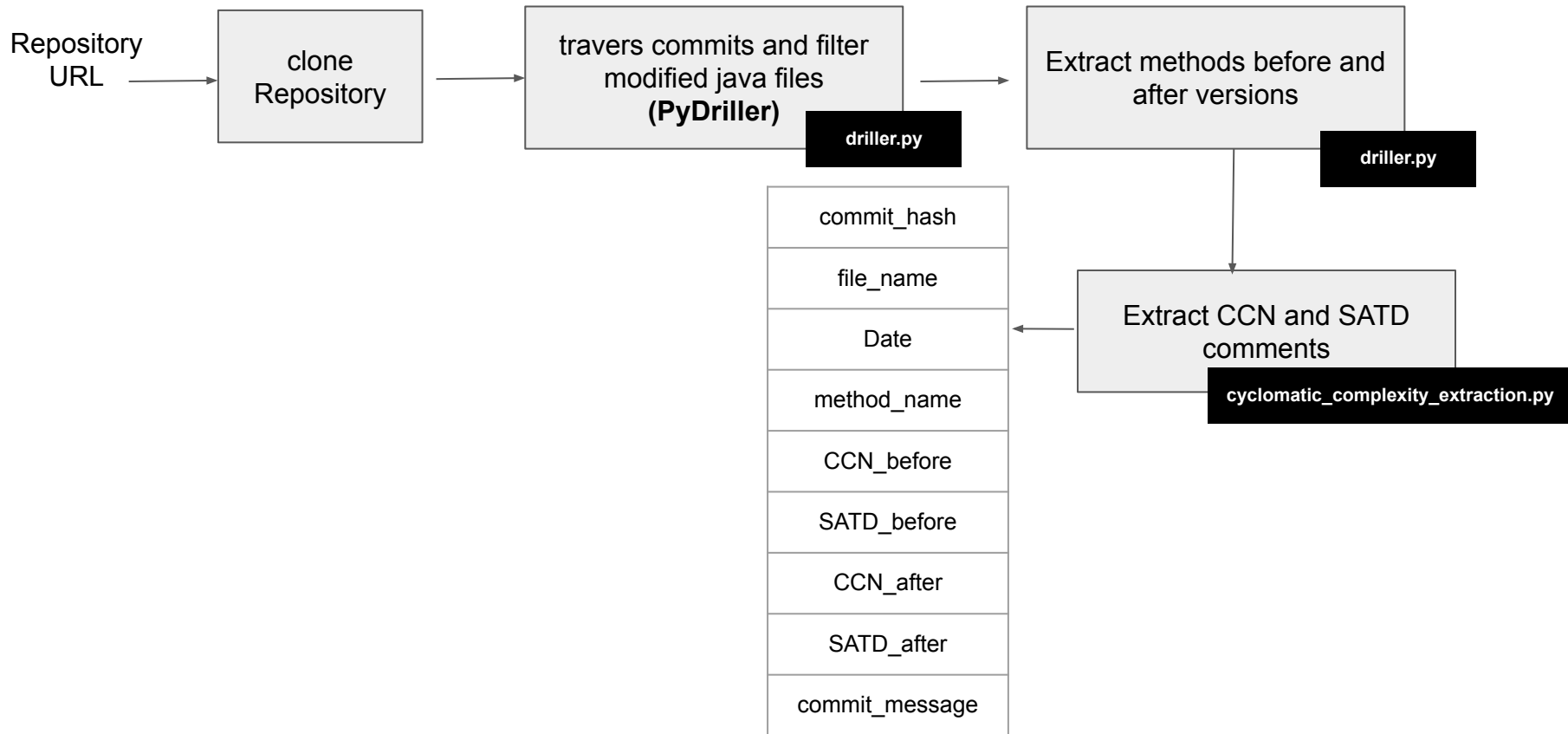
RQ2 Does SATD in commit messages correlate with high code complexity?

Data extraction workflow (1)

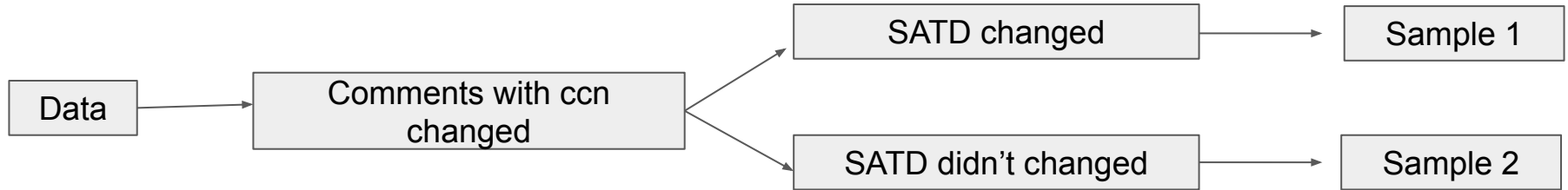


Data extraction workflow (2)





RQ1: Does SATD in code comments correlate with high code complexity?



We added two features `diff_satd` and `diff_ccn`

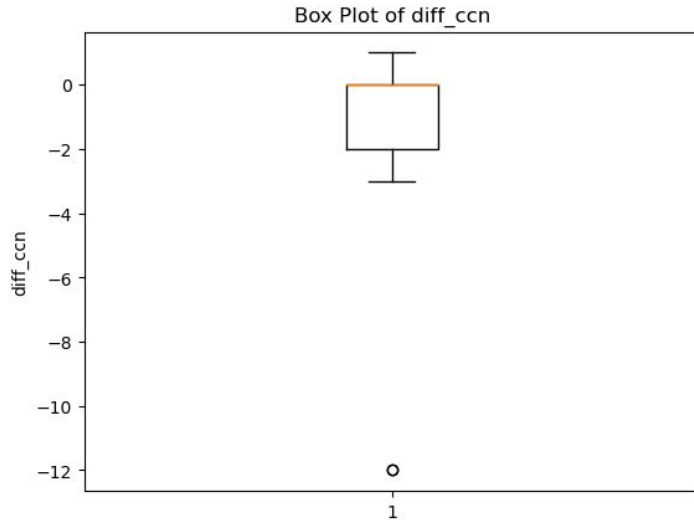
RQ1: Does SATD in code comments correlate with high code complexity?

Comparing samples : hypothesis testing

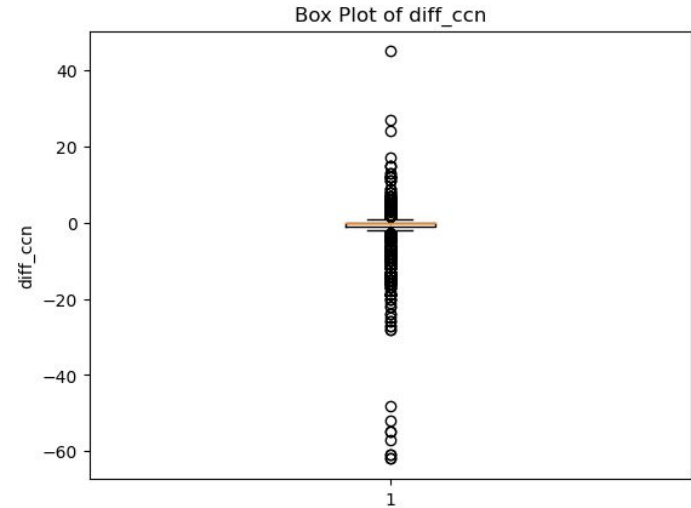
- Significance level :0.05
- H_0 : SATD in comments changes has no effect on the CCN
- H_1/H_a : SATD in comments changes has effect on the CCN

RQ1: Does SATD in code comments correlate with high code complexity?

Code complexity changes distribution : box-plots



Sample 1 (SATDChanged,11)



Sample 2(SATD Not changed, 5322)

RQ1: Does SATD in code comments correlate with high code complexity?

T-test

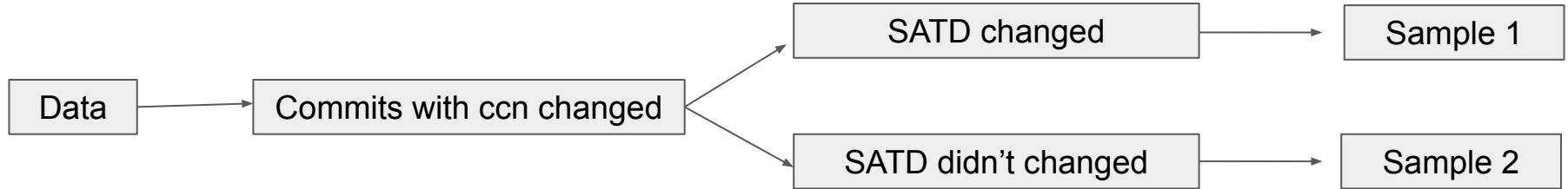
t-statistic: 1.6
p-value: 0.08

Values are not significant we can't
reject or accept the hypothesis

Mann-Whitney U test

U-statistic: 1.3
p-value: 0.16

RQ2 Does SATD in commit messages correlate with high code complexity?



We added a feature : `is_message_satd`

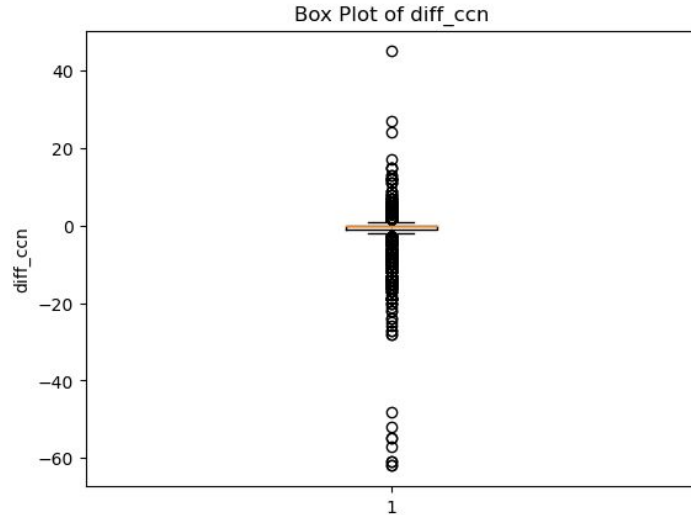
RQ2: Does SATD in commit messages correlate with high code complexity?

Comparing samples : hypothesis testing

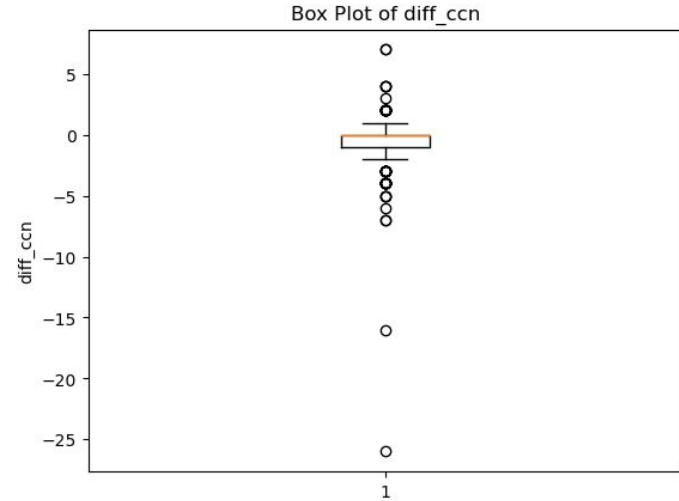
- Significance level :0.05
- H_0 : SATD in commit messages has no effect on the CCN
- H_1/H_a : SATD in commit messages has effect on the CCN

RQ2: Does SATD in commit messages correlate with high code complexity?

Code complexity changes distribution : box-plots



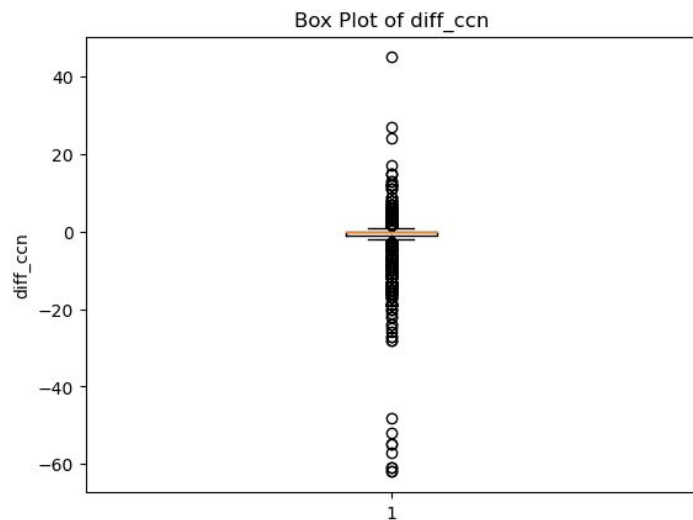
Sample 1(2654) has SATD



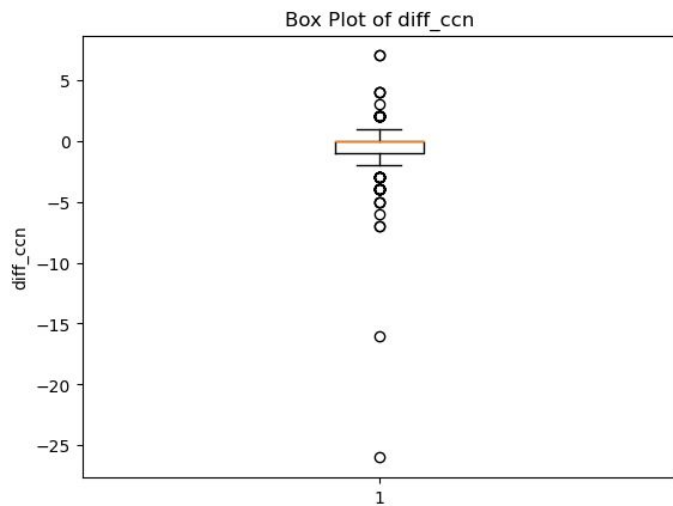
Sample 2 (67) No SATD

RQ2: Does SATD in commit messages correlate with high code complexity?

Code complexity changes distribution : box-plots



Sample 1(2654) has SATD



Sample 2 (67) No SATD

RQ2: Does SATD in commit messages correlate with high code complexity?

T-test

t-statistic: -1.1
p-value: 0.3

Mann-Whitney U test

U-statistic: -0.9
p-value: 0.35

Issues / Further work

- Correlation between specific words like 'Clean up'
 - Better patterns
 - NLP
 - Classification
 - TODO
- Finding repos with significant comments with SATD
- Compare across repository
- Algorithms for SATD patterns detection works with pattern matching, but usually, they tend to have a good precision and a bad recall.

Thank you for your attention