Sentiment Analysis: : **CHEAT SHEET**

About

Kaiaulu supports the download of comments from JIRA, GitHub, Mailing Lists and more. This Cheat Sheet showcases Sentiment analysis, which classifies an author's comment as negative, neutral or positive. Sentiment analysis can be used to assess unhealthy communication patterns in a project, potentially affecting its code quality.

Project Config Setup

The first part of running any vignette is setting up your project configuration file (examples in conf/kaiaulu).

Required Fields

```
• filter:
    replies:
        filter_by_reply_author_substring:
        filter_by_reply_subject_substring:
        filter_by_reply_body_substring:
        regex_to_replace_key_with:
• mailing_list:
    mod_mbox:
        project_key_1:
    pipermail:
        project_key_1:
• issue_tracker:
    jira:
        project_key_1:
            issue_comments:
    github:
        project_key_1:
            issue_or_pr_comment:
• tool:
    sentiment:
        model:
        prediction:
```

You should also specify the required external tools in tools.yml:

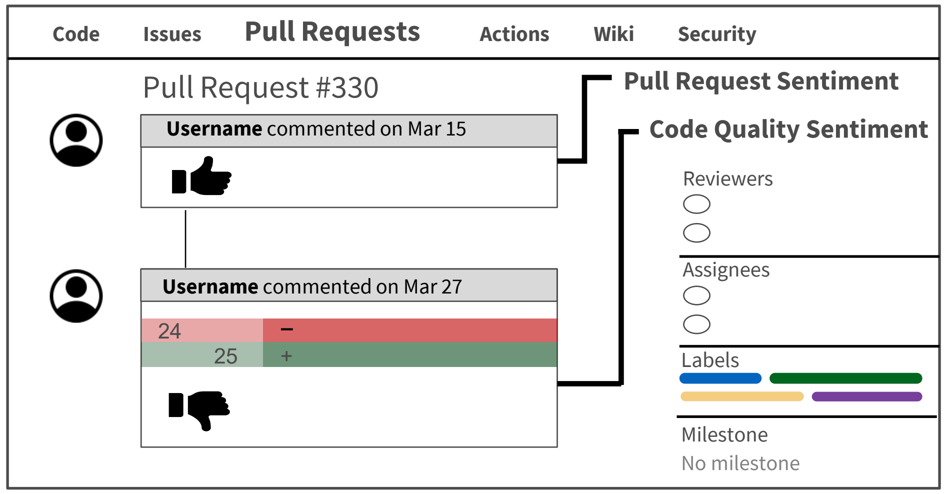
Tools Config Setup

Required Fields

```
• perceval:
• pysenti:
```

Comment's Sentiment

Comments can occur in different data sources (GitHub, Jira, Mailing list), and also occur under different views within the same data source. Sentiment labels carry different meaning depending on the view (e.g. feature communication, specific code line review). Sentiment can also be extracted either from text or emoji.



Pull Request #330

Username commented on Mar 15 👍

Username commented on Mar 27

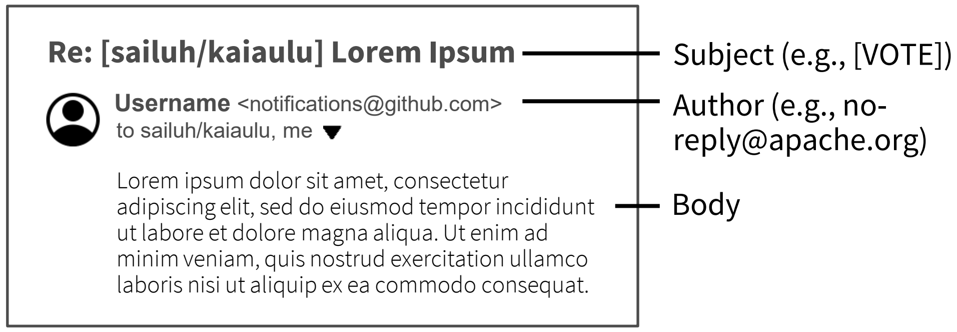
24 −
25 +

👎

Reply Filters

Developer comments generally include code blocks, automated messages, bot messages, markdown formatting, etc. To ensure that the model focuses only on human-generated content, Kaiaulu offers filters to pre-process comment data from any source.

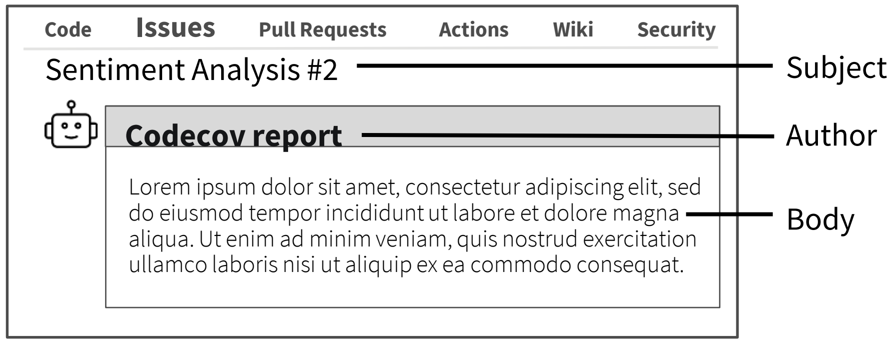
filter_by_reply_<author|subject|body>_substring()



Re: [sailuh/kaiaulu] Lorem Ipsum

Username <notifications@github.com>
to sailuh/kaiaulu, me ▼

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.



Sentiment Analysis #2

Codecov report

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

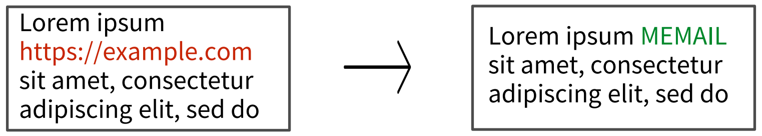
Related Vignettes

See supplementary notebooks that go in-depth on downloading and parsing relevant communication data.

1. **download_mail.Rmd**
2. **download_jira_issues.Rmd**
3. **download_github_issue_comments.Rmd**
4. **download_github_pull_request_comments.Rmd**

replace_token_regex_with()

Replaces text matching given regex patterns with standardized token placeholders (e.g, MURL replaces https://example.com, MEMAIL replaces user@example.com, etc.)



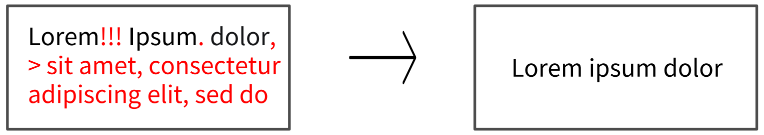
Lorem ipsum https://example.com sit amet, consectetur adipiscing elit, sed do

→

Lorem ipsum MEMAIL sit amet, consectetur adipiscing elit, sed do

filter_text_<punctuation|markdown|newlines|white space|(quoted_lines)|(github_header)>

Text-cleaning functions that remove text content and convert markdown to plain text.



Lorem!!! Ipsum. dolor,
> sit amet, consectetur adipiscing elit, sed do

→

Lorem ipsum dolor

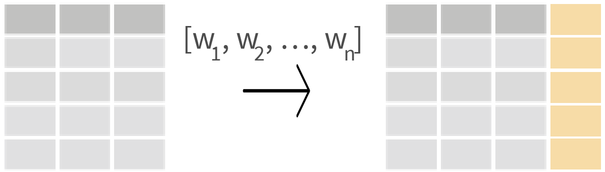
pysenti_train_model(pysenti_path,...)

Trains a specified model using (manually) labeled sentiment data, and saves the weights to the specified project configuration model path.

$[w_1, w_2, \ldots, w_n]$

pysenti_predict(pysenti_path,...)

Loads a pysenti saved model weights to classify the sentiment of a data table's unlabeled comments and returns a data table with comment's classified sentiments.



$[w_1, w_2, \ldots, w_n]$

Kaiāulu

Gerald Huff, Carlos Paradis, Haotian Zhang • Kaiaulu package version 0.0.0.9700 (in development) • Updated: 2025-12