Sentiment Analysis and Kaiāulu React

Austin Dang, Gerald Matthew Huff
Sponsor: Carlos Paradis, Software Analytics Insight Lab
ICS 496 Capstone Project – Spring 2025

github.com/sailuh/kaiaulu

What is Kaiāulu?

Kaiaulu is an R package and common interface that helps with understanding evolving software development communities, and the artifacts (gitlog, mailing list, files, etc.) which developers collaborate and communicate about which allows developers to get a better understanding of the software.

Objectives

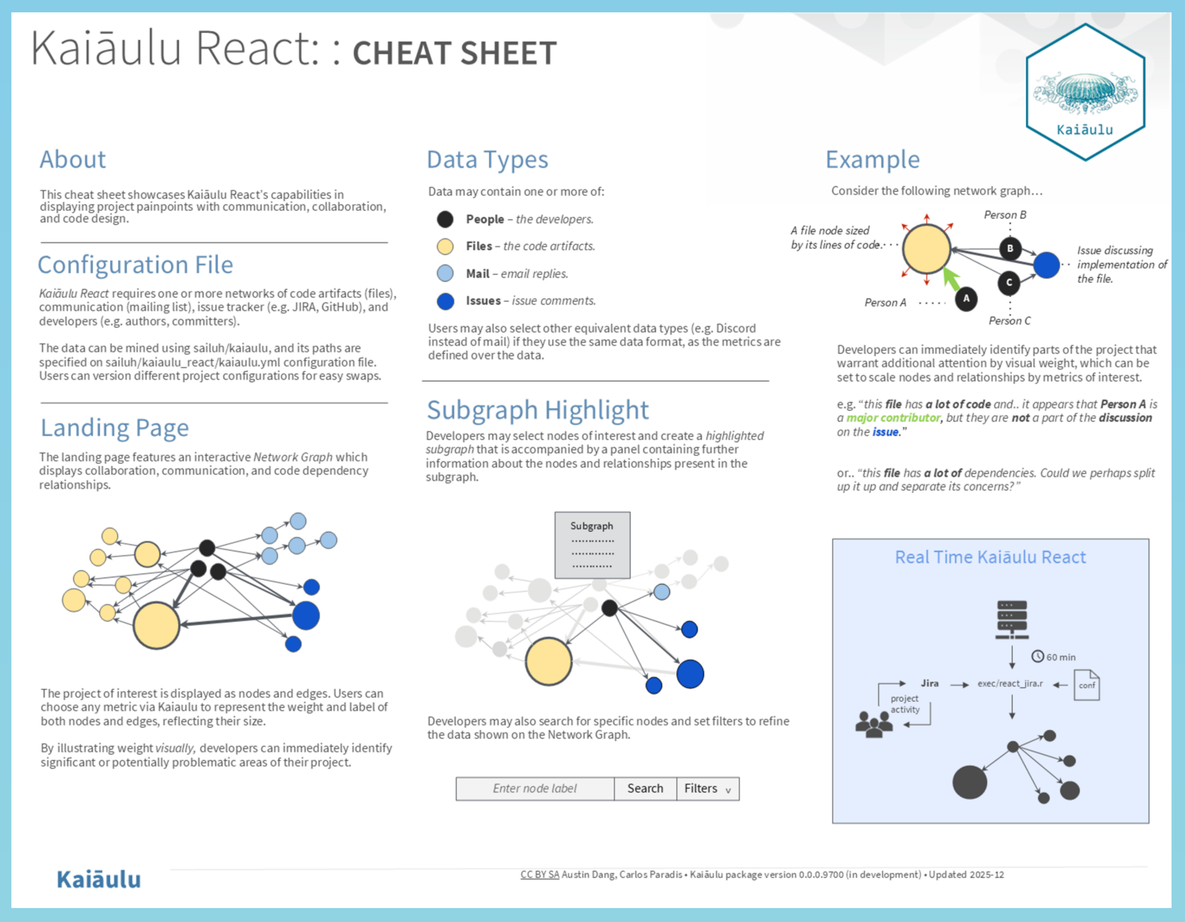
- **Sentiment tools**: Introduce sentiment-analysis tools for project communication data using machine-learning models to help researchers and developers understand team emotional tone and communication trends.
- **Preprocessing improvements**: Improve data preprocessing methods to ensure necessary communication data is consistent and ready for accurate model training and sentiment prediction.
- **Web application for software analytics:** Create a web application for software analysis by providing an interactive interface for visualizing data extracted by Kaiaulu.
- **Scalable web architecture:** Establish a scalable architecture that supports the efforts of future developers in building upon the Kaiaulu web application.

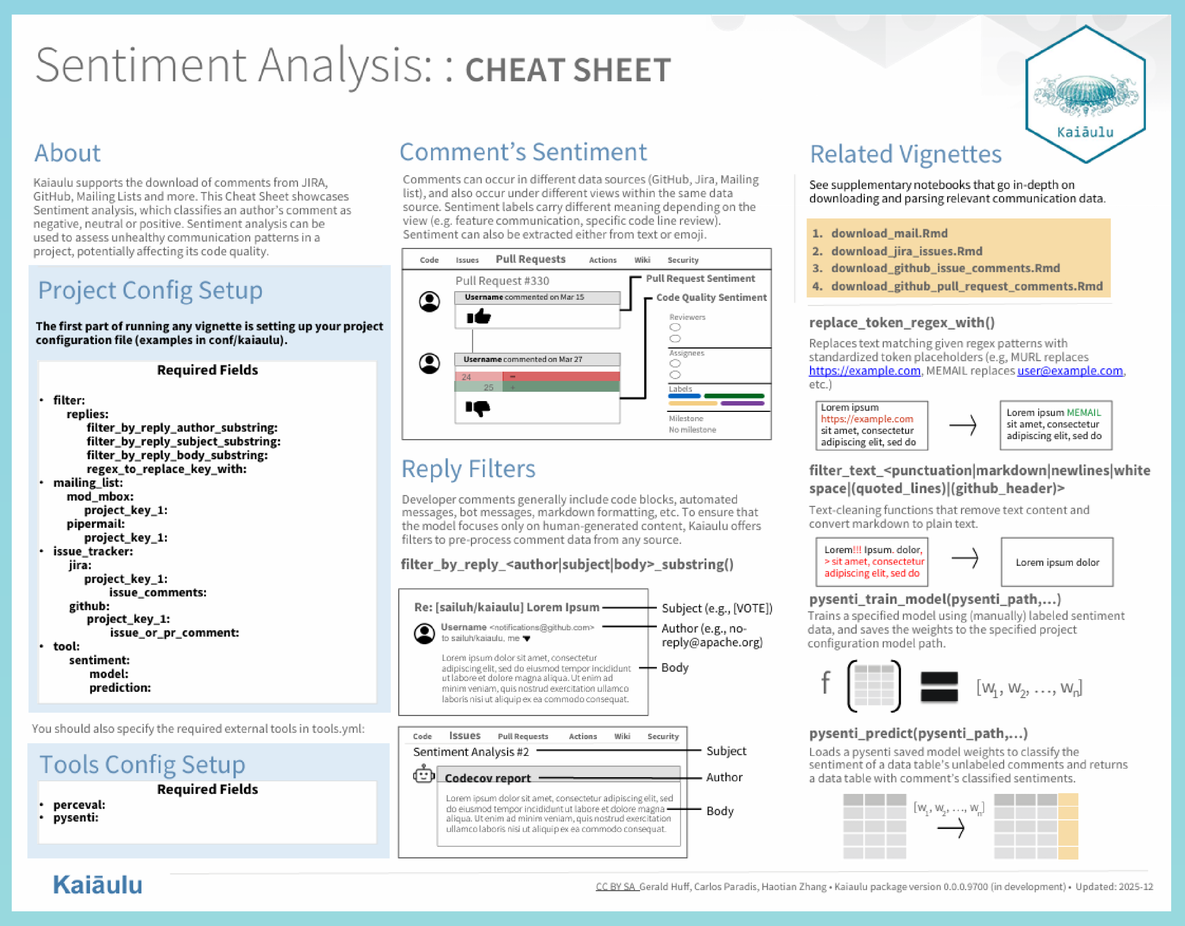
Challenges

- Understanding diverse communication data was challenging; emails, comments, and issue trackers. store information differently, requiring tailored preprocessing to filter out irrelevant text so the ML model can focus on meaningful content.
- Implementing a consistent architecture when adding new features to the existing Kaiaulu repository, requiring careful adherence to the established workflows and design patterns.
- Designing a robust architecture for building a web application that will be understood and expanded upon by future contributors.
- Utilizing a JavaScript data visualization library to create a compelling interface for software analysis that supports the capabilities of the Kaiaulu R package.

Tasks accomplished

- Built a preprocessing pipeline for cleaning and standardizing communication data (emails, Github comments, Jira issues)
- Configured project workflows using standardized YAML files for reproducibility.
- Integrated Python-based ML scripts into R notebooks to perform sentiment analysis, creating a reusable method that enables future users to assess overall sentiment in new datasets and better understand developer communication trends.
- Created a web application that utilizes data extracted by the Kaiaulu R package and provides several features to assist developers in analyzing software.
- Created documentation that will inform new contributors of the logic behind existing features and how to expand upon the application while maintaining proper organization.





Technologies

- Git & GitHub
- Languages: TypeScript + CSS, Python, R
- JavaScript Libraries: React, D3.js, MUI
- Third Party Tools: Perceval, Pysenti

Learnings & Conclusions

The project reinforced the importance of adhering to Kaiaulu's architecture. By implementing reproducibility methods within preprocessing steps, it can be adapted to different communication sources which supports flexible sentiment-analysis workflows that can be reliably applied across projects and makes the overall process easier to manage.

Creation of the Kaiāulu web application from the ground up allowed for a strong reinforcement of development skills in addition to a newfound understanding of architectural design. The rippling effects of design choices made at the beginning of this project was an invaluable learning experience. Furthermore, deep insight was gained when designing an application with developers in mind and even more regarding data visualization.