Team Name: ByteCraft

Members:

- Ben Sivoravong (bs62, bs62@illinois.edu)
- Shivani Mangaleswaran (sm131, sm131@illinois.edu)
- Yogi Patel (ypatel55, ypatel55@illinois.edu)
- *Captain*: Annamika Dua (ad8, ad8@illinois.edu)

Topic: Automatic issue triage for Github issues

Progress Report:

1) Which tasks have been completed?

- Initial investigation of different python libraries. Deciding which modules, APIs we want to use for our application:
    - ***nltk:***
        - https://www.nltk.org
        - Used for tokenization, parsing, classification, stemming, tagging, and semantic reasoning
            - We can use it in our project for tokenization, filtering stop words, POS tagging
            - https://realpython.com/nltk-nlp-python/#getting-started-with-pythons-nltk
    - ***sklearn:***
        - We can use it to create a BoW vector representation of the words in the issues, and use TF-IDF ranking to rank the maintainers
        - https://pages.github.rpi.edu/kuruzj/website_introml_rpi/notebooks/08-intro-nlp/03-scikit-learn-text.html
    - ***pandas*** + ***numpy***: can be used but may not need
        - Can be used for analyzing, cleaning, and exploring data
        - https://www.dataquest.io/blog/how-to-clean-and-prepare-your-data-for-analysis/
    - tensorflow (word2vec): will not be used since it doesn't seem necessary for our use case
        - CBOW (Continuous Bag of Words): CBOW model predicts the current word given context words within a specific window
        - https://www.geeksforgeeks.org/python-word-embedding-using-word2vec/
- Overall system design, technical architecture:
    - See "docs" folder in repo for design and architecture.

- Start coding to connect with Github REST APIs, as well as tokenize the responses: https://docs.github.com/en/rest?apiVersion=2022-11-28
    - Read issue text
    - Get contributors for a project
    - Webhooks, trigger our lambda or something on new issues

2) Which tasks are pending?
- Further implementation and testing of the application
    - We will first test the local implementation, and then move onto the Github webhooks implementation

3) Are you facing any challenges at the moment?
- We may face some challenges as we continue on with the implementation of our project. For example, one possible challenge is a lack of maintainers in a given repo. There are some OSS repos like JOOQ, where usually only one person handles most issues. So if there are few active maintainers, then we are more likely to assign most issues to only that one person.