



Prompt Analysis : ChatGPT

Abhijeet Shah, Shahzeb Faheem, Tianyun Liu, Wonjong
Kim, Yunpeng Wen

Research Question

- I. Can we reliably predict whether a developer's issue will be resolved based on the initial conversation with **ChatGPT**?
- II. If developers were to rerun their prompts with **ChatGPT** now and/or with different settings, would they obtain the same results?



Motivation

- I. With the rise of ChatGPT, developers are shifting from StackOverflow
- II. ChatGPT provides a real time discussion based experience as opposed to StackOverflow.
- III. It is imperative to know how to manipulate ChatGPT in order to get successful answers.





What we have done so far

- I. Extracted the data from the DevGPT compiled developer-centric issues (988 Conversation Threads).
- II. Classified the data based on resolved and unresolved categories.

How have we classified the data?

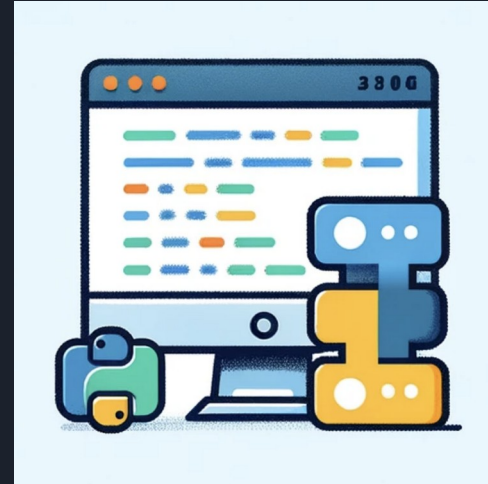
```
"Sources": [  
  {  
    "Type": "issue",  
    "URL": "https://github.com/gakusyutai/gakusyutai.github.io/issues/31",  
    "Author": "yuyu31",  
    "RepoName": "gakusyutai/gakusyutai.github.io",  
    "RepoLanguage": "HTML",  
    "Number": 31,  
    "Title": "\u30cf\u30f3\u30d0\u30fc\u30ac\u30fc\u30e1\u30cb\u30e5\u30fc\u306e\u5b9f\u88c5",  
    "Body": "- https://chat.openai.com/share/8b0f517f-1aaf-4b50-8d79-1138941a383f\r\n- https://chat.openai.com/share/7958273c-0081-48d1-8d15-0dd8a5944457\r\nChat GPT \u3092\u4f7f\u3063\u3066\u5b9f\u88c5\u3092\u8a66\u307f\u3066\u3044\u308b",  
    "CreatedAt": "2023-07-23T15:38:42Z",  
    "ClosedAt": null,  
    "UpdatedAt": "2023-07-23T15:38:42Z",  
    "State": "OPEN",  
    "ChatgptSharing": [  
      {  
        "URL": "https://chat.openai.com/share/7958273c-0081-48d1-8d15-0dd8a5944457",  
        "Mention": {  
          "MentionedURL": "https://github.com/gakusyutai/gakusyutai.github.io/issues/31",  
          "MentionedProperty": "body",  
          "MentionedAuthor": "yuyu31",  
          "MentionedText": "- https://chat.openai.com/share/8b0f517f-1aaf-4b50-8d79-1138941a383f\r\n- https://chat.openai.com/share/7958273c-0081-48d1-8d15-0dd8a5944457\r\nChat GPT \u3092\u4f7f\u3063\u3066\u5b9f\u88c5\u3092\u8a66\u307f\u3066\u3044\u308b"  
        },  
        "Status": 200,  
        "DateOfConversation": "July 6, 2023",  
        "DateOfAccess": "2023-07-27 13:18:40.616483",  
        "Title": "\u30cf\u30f3\u30d0\u30fc\u30ac\u30fc\u30e1\u30cb\u30e5\u30fc\u8868\u793a\u306e\u4fee\u6b63",  
        "NumberOfPrompts": 11,  
        "TokensOfPrompts": 2637,  
        "TokensOfAnswers": 4712,  
        "Model": "Default",  
        "Conversations": [  
          {  
            "Prompt":  
              "\u3042\u306a\u305f\u306fweb\u30c7\u30b6\u30a4\u30ca\u30fc\u3067\u3059\u3002\u30cf\u30f3\u30d0\u30fc\u30ac\u30fc\u30e1\u30cb\u30e5\u30fc\u3092\u5b9f\u88c5\u3057\u305f\u3068\u3053\u308d\u3001\u521d\u3081\u304b\u3089\u30e1\u30cb\u30e5\u30fc\u306e\u5185\u5bb9\u304c\u8868\u793a\u3055\u308c\u3066\u3044\u3066\u3001\u8868\u793a\u975e\u8868\u793a\u3092\u5207\u308a\u66ff\u3048\u3089\u308c\u307e\u305b\u3093\u3002\u4ee5\u4e0b\u306e\u3088\u3046\u306a\u30b3\u30fc\u3092\u7528\u610f\u3057\u3066\u3044\u308b\u3068\u304d\u3001\u3069\u306e\u3088\u3046\u306a\u4fee\u6b63\u304c\u800
```

Design and Evaluation

DESIGN

I. Use of ML Classifier

- Create a Machine learning classifier with a baseline of 50% and train it on a dataset of 395 solved + 395 unsolved queries (80% of dataset) between the developer and ChatGPT.

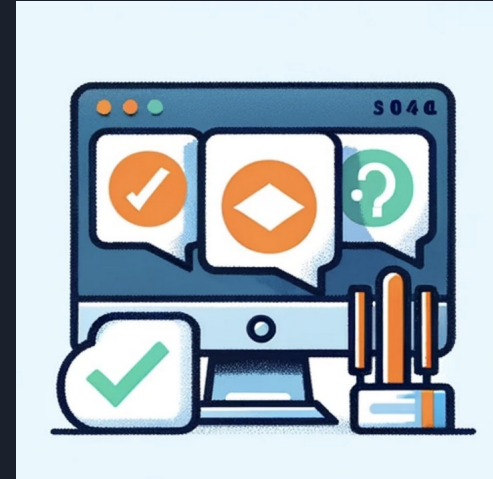


Proposed Design and Evaluation

DESIGN

II. Model Selection

- Random Forest classifier



Proposed Design and Evaluation

DESIGN

III. Feature Engineering

- Use TF-IDF vectorization for text features. It gives a score to each keyword on the number of times it appears in the text.



Proposed Design and Evaluation

DESIGN

IV. ChatGPT API

- Test the prompts from the issues dataset using ChatGPT API under various different settings like varying temperature, max tokens.





Proposed Design and Evaluation

EVALUATION

I. Metrics Evaluation

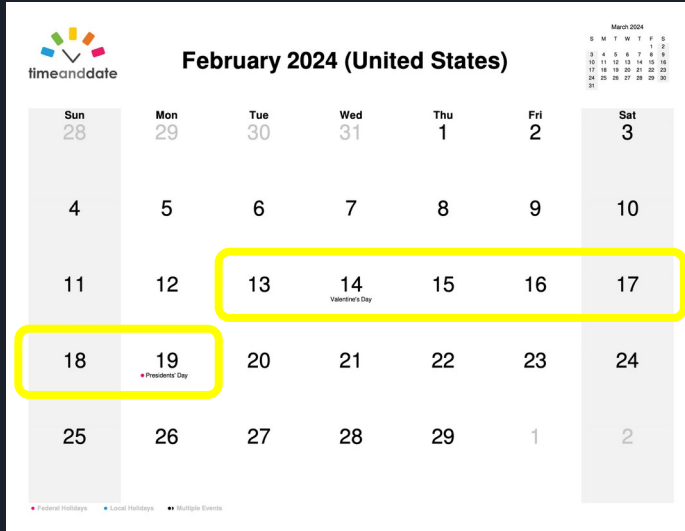
- Check for accuracy of prediction, precision, True Positive Rate.
(First research question)

II. Statistical Analysis

- Conducting Paired T-Test, i.e., Assess the significance of differences between responses by comparing the mean scores of paired samples calculated by cosine similarity.
(Second research question)

Development Phase

TIMELINE

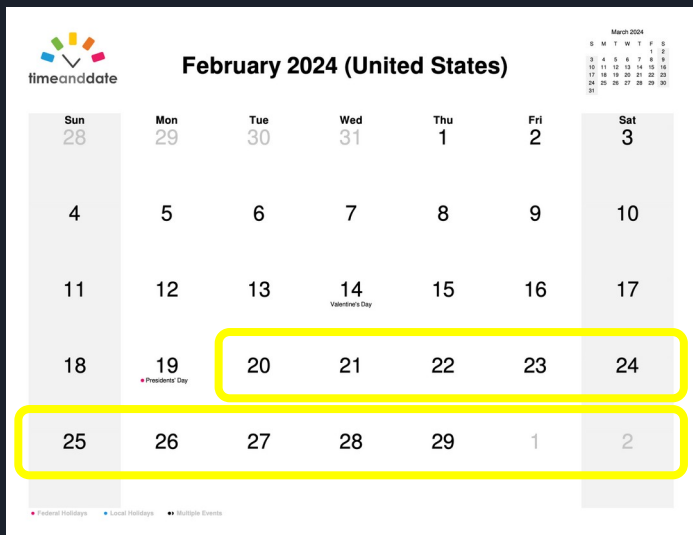


Develop ML Model

- Select the appropriate ML Model (Random Forest Classifier) and train it on a dataset of 395 solved + 395 unsolved issues conversation threads.

Evaluation Phase

TIMELINE



Testing the Trained ML Model

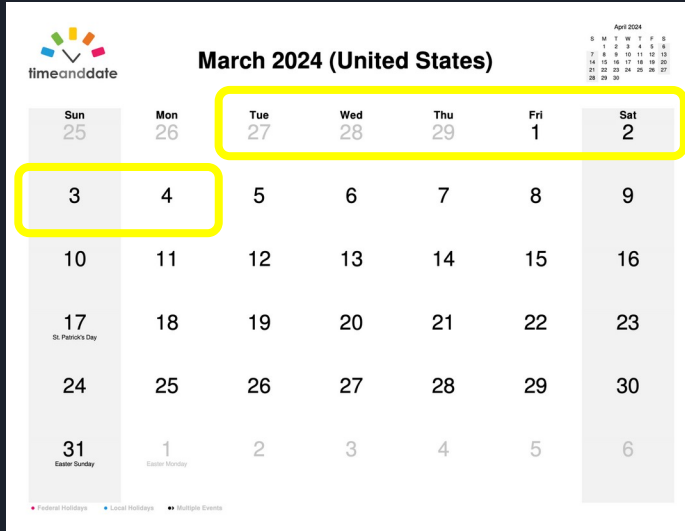
- Run the trained ML Model on the complete dataset and predict the accuracy.

ChatGPT API

- Test ChatGPT API on different settings and initiate statistical analysis (Paired t-Test) on the old vs new answers.

Testing Phase

TIMELINE

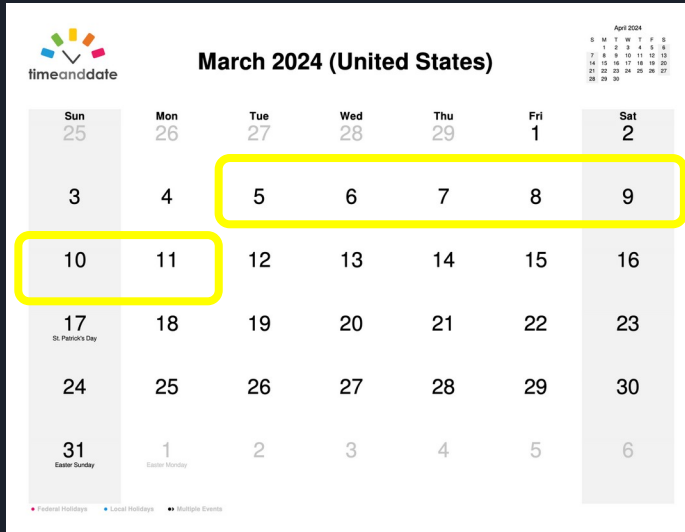


Report

- Compile findings into a comprehensive report and discuss and refine conclusions.

Conclusion

TIMELINE



Presentation Preparation

- Prepare slides for the Final Project and rehearse.



Responsibilities

Shahzeb - Lead Developer RQ2 | Presentation Lead

Wonjong - Support RQ2 | Report Lead.

Tianyun - Lead Statistical Analyst | Support Role : Report and Presentation.

Abhijeet - Data Miner | Support Developer RQ1 | Project Manager

Yunpeng - Lead Developer ML Model | Support Role : Report.



THANK YOU