ML - Assignment

Problem Statement: Channel Agnostic Summary Generation for Customer Feedback

Given a feedback record, generate a useful summary of the user's intent. The intents we're interested in are Complaint, Help, Improvement or Praise.

```
- *Complaint*: Indicates a problem or challenge faced by the user.
- *Praise*: Reflects appreciation or positive feedback.
- *Improvement*: Suggests a need for enhancement or modification.
- *Help*: Signals a request for assistance, guidance, or support.
- *None*: Does not fit into any of the above categories.
```

Make sure to capture the user's dialouge in great detail so that as a user when i read the summary, i have a crisp and clear understanding of what was being talked about in the feedback record.

`None` is a valid summary for when no substantial useful information is present in the feedback.

Dataset:

You are provided with a Dataset of 5010 records and their summaries. Develop a model, that can generate summaries for any kind of records.

Link to Dataset: https://docs.google.com/spreadsheets/d/1MyMeEkZ-mFQRNwiEu_LBy4CjMhZzJc_Xxipx_ez-doU/edit?usp=sharing

There are 3 Data-sources in the dataset

- RecordTypeReview - From Playstore/Appstore

- RecordTypeForumConversation - Twitter (https://twitter.com/)

- RecordTypeSurvey - G2 (https://www.g2.com/)

Annotation Guidelines for the Summaries:

We provide you, with the annotation guidelines that our annotators have used to write summaries for records.

The Annotation Guidelines should help formulate an idea of what the ideal summary for a record looks like. Make sure you go through the entire contents of the Annotation Guidelines properly as it dictates what is expected in a summary across a variety of Record Types.

1. Read the entire content carefully: Before deciding whether to create a summary, ensure that you have read and understood the whole review, survey, support conversation, app store/play store review, NPS, or user recording to get a clear grasp of the user's concerns, expectations, and feedback.

2. Evaluate the usefulness of the content: Determine if the content provides meaningful information or feedback that would be valuable to the product team. If the content lacks substantial information or is just a conversation with a happy resolution and no actionable insights, the summary should be "None." If the content has some information but is not detailed enough to be actionable, try to summarise the general issue or concern without assuming additional details.

3. Identify the "What": If the content is deemed valuable, determine the main subject of the content, which is usually the product, service, or specific feature being discussed. Summarise the user's experience or opinion about it in a few words.

4. Address the "Why": Explain briefly why the user has the opinion they expressed in the content. This may include any specific reasons, challenges, or issues they faced while using the product or service. If the "Why" is not explicitly mentioned but can be reasonably inferred, include it in the summary.

5. Mention the "How": If the content provides any solutions or suggestions, include them in your summary. This may involve the user's recommendations, steps they took to resolve an issue, or alternative methods they tried.

6. Keep it concise: The summary should be short, yet informative. Aim to capture the essence of the content in one or two sentences, without using unnecessary words or repeating information.

7. Use clear and simple language: Write the summary using clear and straightforward language, avoiding jargon or complex terms. Make sure it is easily understandable by a wide audience.

8. Maintain objectivity: Ensure that your summary is unbiased and does not include your personal opinions.

9. Adapt to the content type: While the general guidelines remain the same, be prepared to adapt your summary to the specific type of content you are summarising. For example, support conversation tickets may require a focus on the resolution and steps taken, while app store/play store reviews may require highlighting the user's overall satisfaction and any standout features.

10. Highlight key insights: In case of surveys or NPS, emphasize the key insights or trends that emerge from the user's responses. For user recordings, focus on the most important takeaways from the user's interaction with the product or service.

11. Ensure accuracy: Make sure your summary accurately represents the user's experience and opinions, without distorting or exaggerating their feedback.

12. Focus on the user's issue: When summarising support conversations, prfioritize the user's problem or concern, rather than the agent's actions, to ensure the summary is useful for understanding the user's issue.

13. Maintain a consistent narrative: Use a consistent narrative style throughout the summary, regardless of the feedback source. Start with the user's issue or opinion, followed by the reasons or challenges they faced, and end with any solutions or suggestions provided. Avoid using questions or statements that do not convey a clear understanding of the user's issue or opinion.

14. Capture implied information: If the user's feedback implies a reason for their opinion or issue, try to include it in the summary, even if it is not explicitly stated. This will help provide a more comprehensive understanding of the user's experience.

15. While summarising a RecordTypeForumConversationThread (Twitter) fetch the last sent message , use the context of the message to summarise the last message, don't mention any personal information like username , twitter handle in the summary .

```
16. Understand the questions asked by Agent and Summarise the User Response Accordingl
y ,use the Agent Question to draw context fot the answers.
```

Expectations

- We expect the candidate to spend no more than 3 days on this assignment.

- We expect, by end of the assignment, for the candidate to have a clear understanding of the problem statement and the challenges involved in scaling the model. Bonus points for identifying gaps in annotation guidelines, dataset summaries that could cause confusion in the model and it's training.

- Develop a model, of your choice, for generating summaries for records. The model will be used to infer on a vast variety of RecordTypes.

  - Additionally, Candidates are expected to wrap their model behind an API endpoint, that can be hit, with text, product name and record type to infer the summary.

- Suggest strategies to evaluate the model for quality of summary. Be creative, we expect you to think about what kind of metric would be a close leading indicator of the quality of summary. Where quality is subjective to our use-case of summarising user feedback in Records.

- A thorough error analysis of the model and data, along with points on how would you go about improving the model further is expected.

Output

- Code: Code for training, data analysis and API

- Report: A log of the thought process, experiments conducted and conclusions drawn from them, followed by Error Analysis of the model.

- API: An API end point, that can be hit, to generate the output from concerned model.