# Executive Summary: Phase 2

## Part 2: Implementation of Zero-Shot and N-Shot Learning for Comment Classification

### Objective

The primary goal of this phase was to implement and evaluate advanced machine learning techniques—specifically zero-shot and N-shot learning—for automating the classification of comments in collaborative documents. Additionally, we explored topic modeling using LDA (Latent Dirichlet Allocation) to understand the underlying themes within the comments. These methods were assessed for their effectiveness in categorizing comments with minimal labeled data.

### Methodology

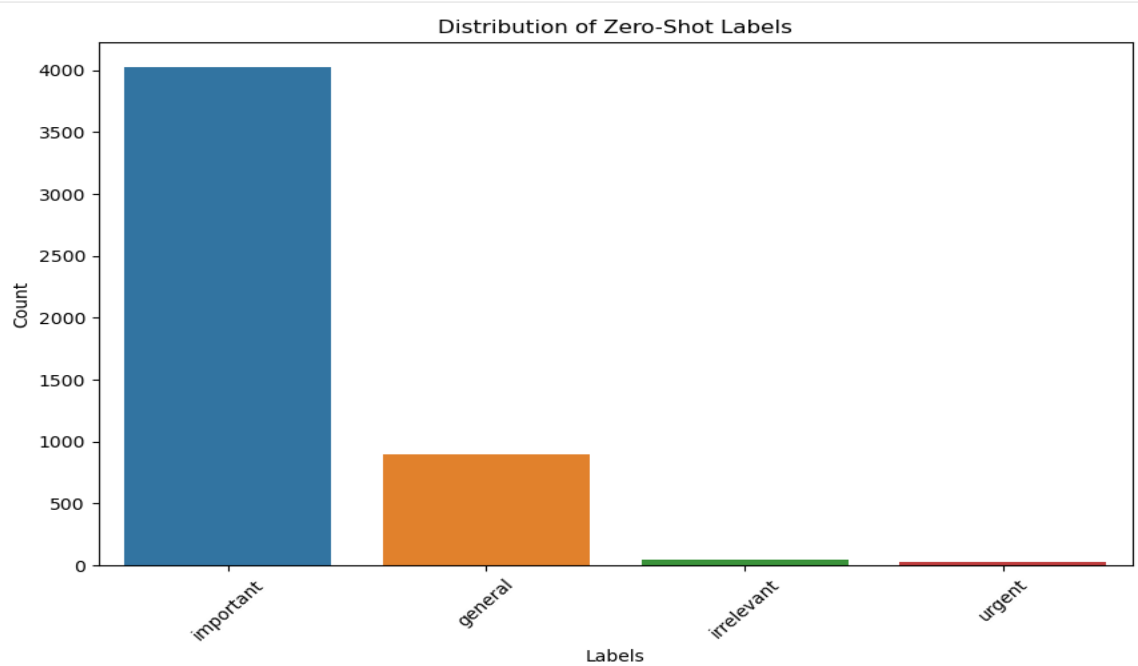
### Zero-Shot Learning:

**Approach:** We utilized the BART-large model for zero-shot classification. This method required no prior training on the specific dataset, leveraging the model's pre-trained capabilities to assign comments to predefined categories such as "INFORMATION EXCHANGE," "MODIFICATION," and "SOCIAL COMMUNICATION."

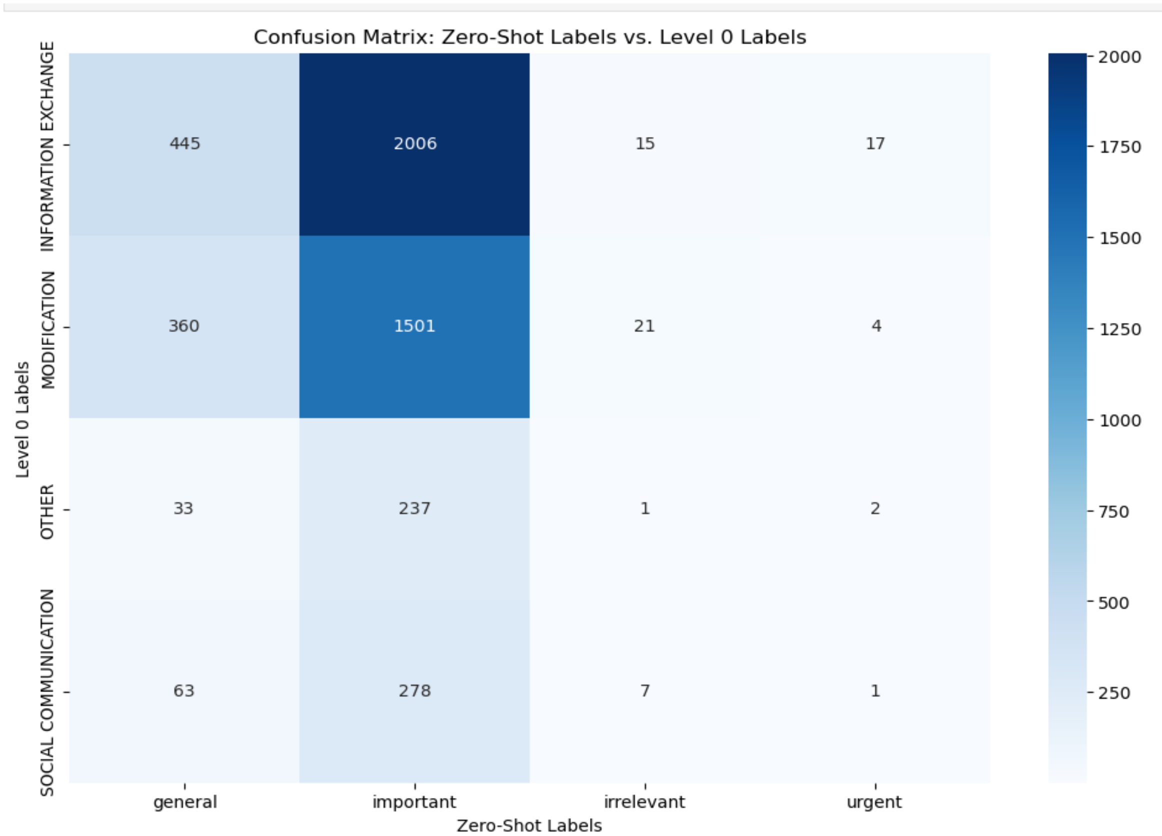
**Evaluation:** The model's predictions were compared against actual labels using metrics of accuracy, precision, recall, and F1 score.

**Visualization:**

- We analyzed the distribution of zero-shot labels across the dataset, visualizing the frequency with which each label was assigned.



- A confusion matrix was generated to compare the zero-shot predictions against the true labels, helping us identify areas where the model performed well and struggled.



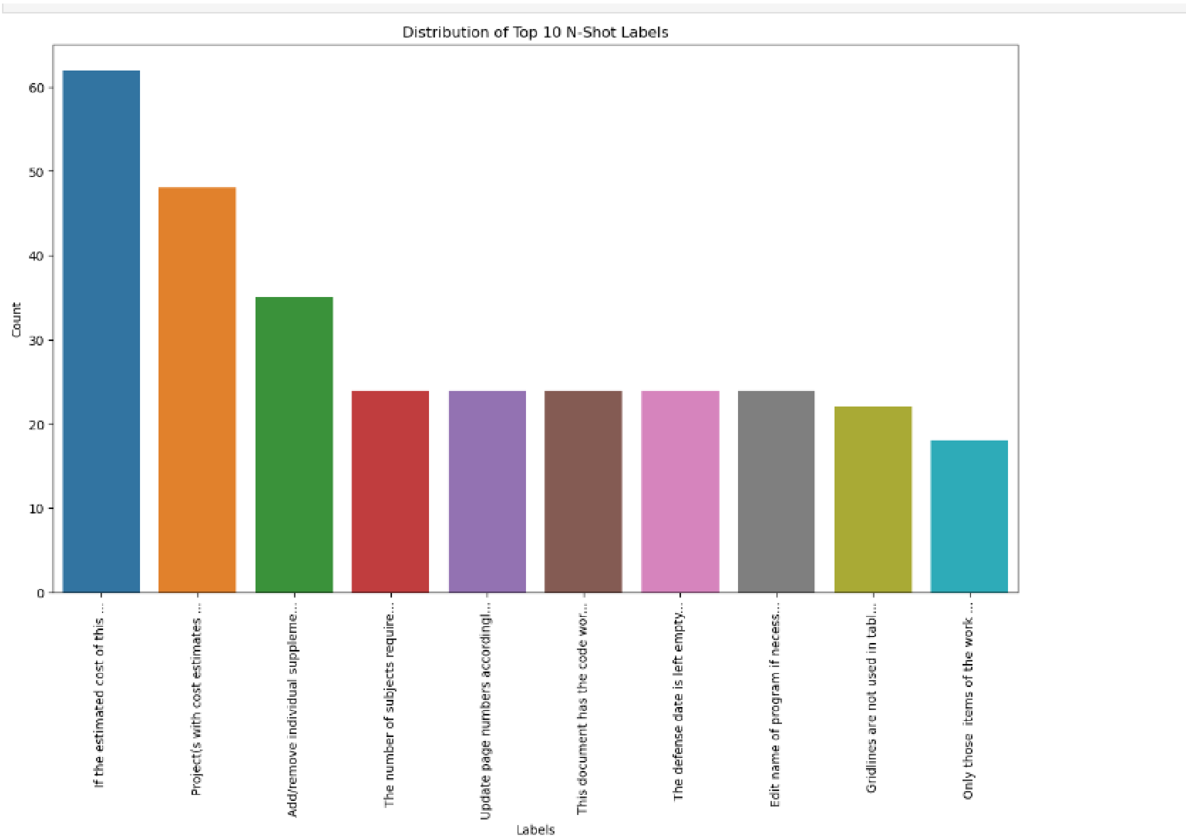

**N-Shot Learning:**

**Approach:** The DistilBERT model was fine-tuned using a small number of labeled comments (N-shot learning). This allowed the model to improve its classification performance with limited training data.

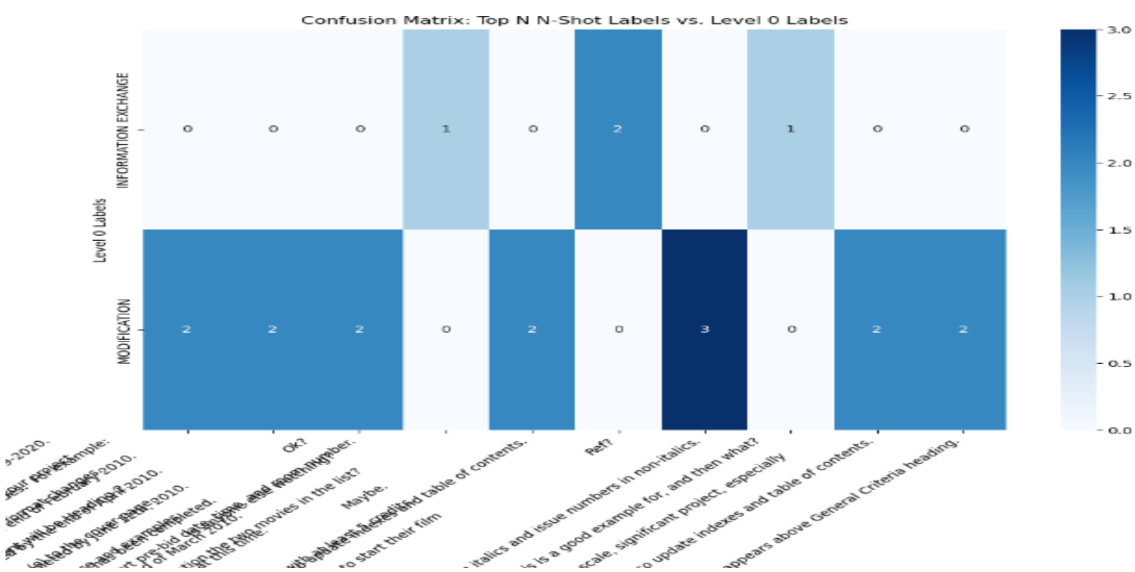
**Evaluation:** The model's performance was tracked over several epochs, with improvements in accuracy, precision, recall, and F1 score.

**Visualization:**

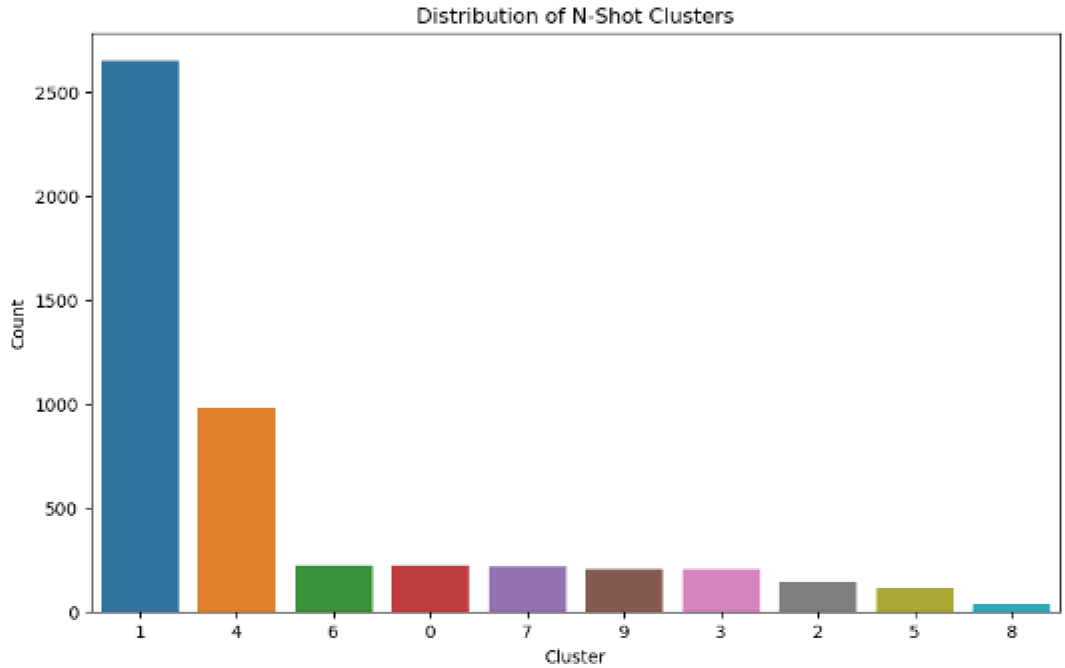
- The distribution of the top N-shot labels was visualized to assess how the model categorized the comments after fine-tuning.



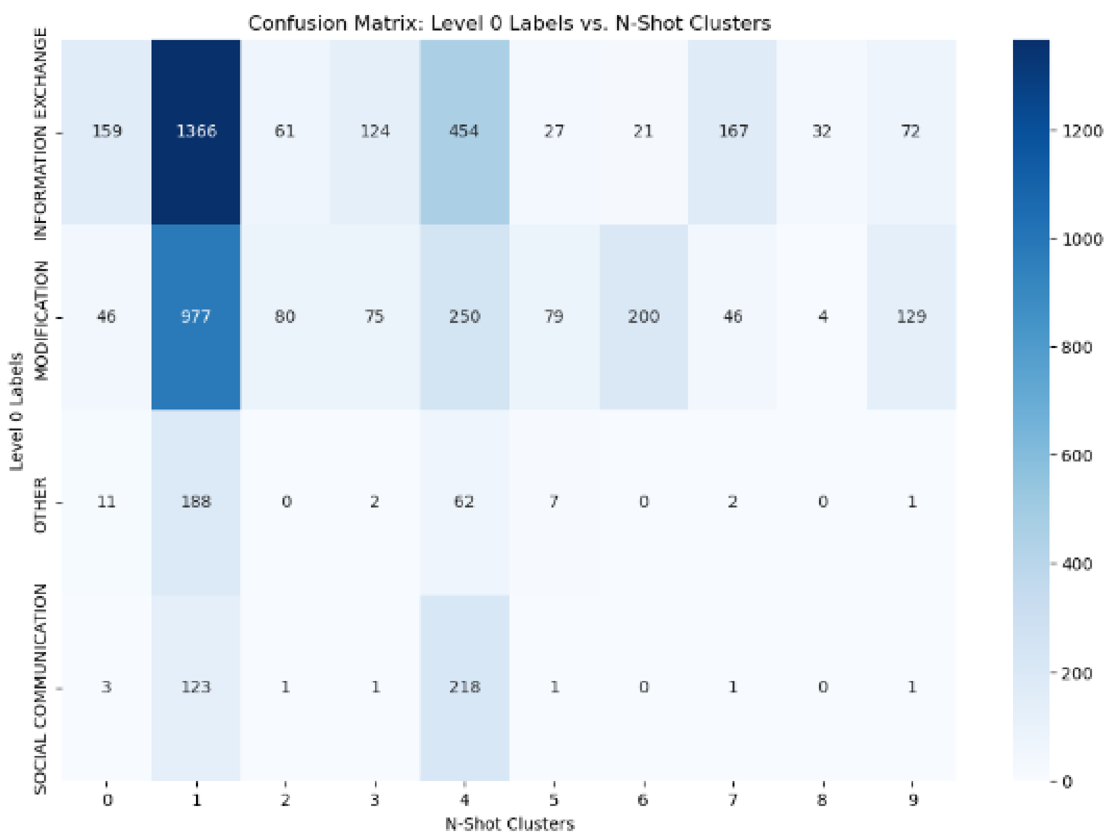

- A confusion matrix comparing N-shot predictions to the actual labels provided insight into the model's effectiveness.




Since the data predicted by n shot learning is very lengthy we grouped them into clusters and below is the visualizations of both bar graph and confusion matrix.



Distribution of N-Shot Clusters
Count
2500
2000
1500
1000
500
0
1
4
6
0
7
9
3
2
5
8
Cluster

Confusion Matrix: Level 0 Labels vs. N-Shot Clusters
Level 0 Labels
INFORMATION EXCHANGE
MODIFICATION
OTHER
SOCIAL COMMUNICATION
159 1366 61 124 454 27 21 167 32 72
46 977 80 75 250 79 200 46 4 129
11 188 0 2 62 7 0 2 0 1
3 123 1 1 218 1 0 1 0 1
0 1 2 3 4 5 6 7 8 9
N-Shot Clusters
1200
1000
800
600
400
200
0

After doing cluster analysis there is a method called lda topic modelling came into our knowledge and that is how we used LDA topic modelling in order to understand indepth hidden pattern of each topic.

**LDA Topic Modeling:**

**Clustering Approach:** We employed clustering techniques to uncover natural groupings within the comments. This helped us identify patterns and themes without predefined categories, making understanding the main topics discussed in the dataset easier.

**What Clustering Showed:**

**Main Themes:** The clustering process revealed that many comments focused on general operational details, such as instructions, guidelines, and project updates.

**Specialized Topics:** Smaller clusters highlighted more specific discussions, such as legal or contractual details and technical research topics.

**LDA Topic Modeling:** LDA was used to dig deeper into the topics identified through clustering. This technique allowed us to identify specific themes within the groups and understand the distribution of these themes across different clusters.

**What LDA Analysis Suggested:**

**Detailed Topics:** LDA identified vital topics such as project updates, research details, and document management. Each topic was characterized by keywords, providing a clear picture of the primary discussions within the comments.

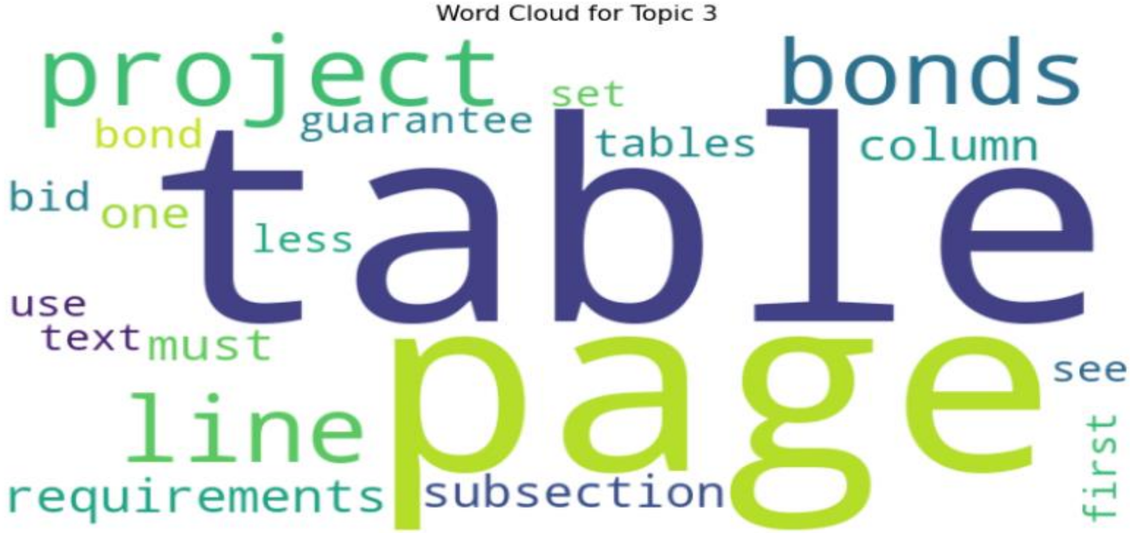
**Sub-Themes:** Within the broader topics, LDA revealed sub-themes such as meeting schedules and research progress, offering more profound insights into the content of the discussions.

**Visualization:**

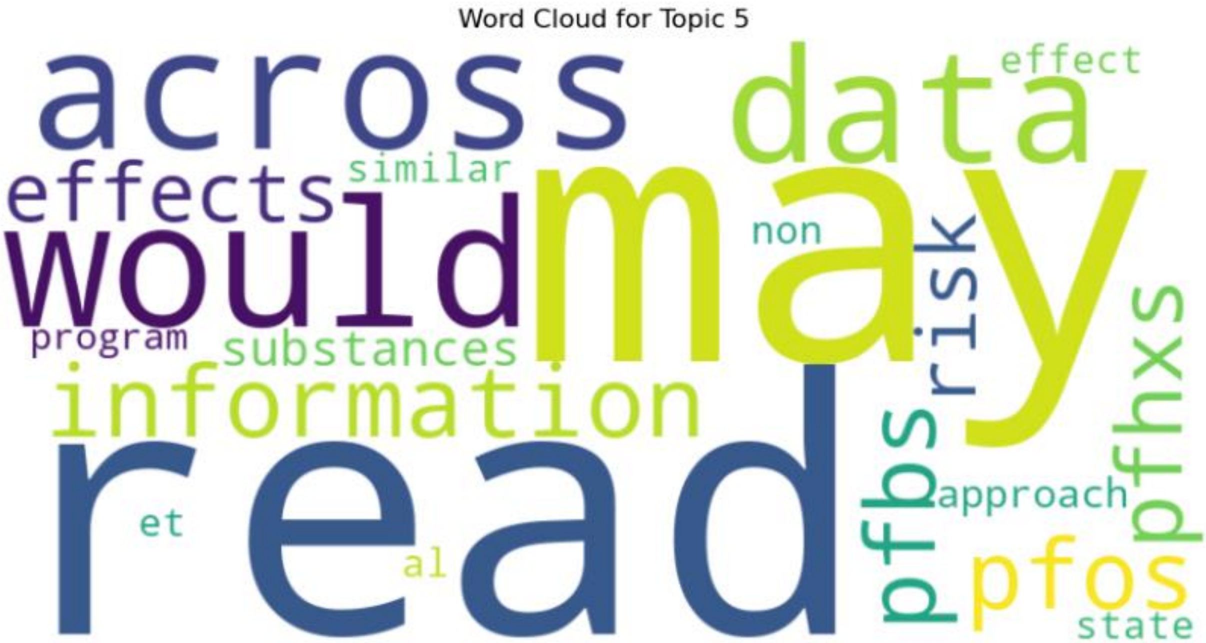
- Word clouds were generated for each topic to visually represent the most prominent terms within the topics identified by LDA.

Word Cloud for Topic 2

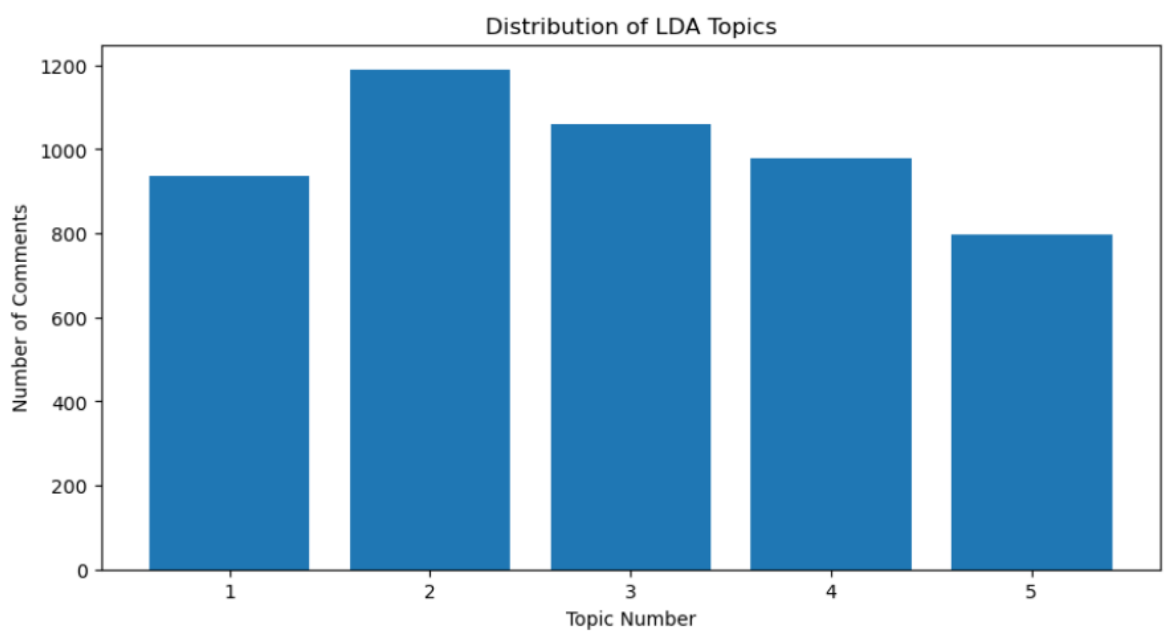

Word Cloud for Topic 3



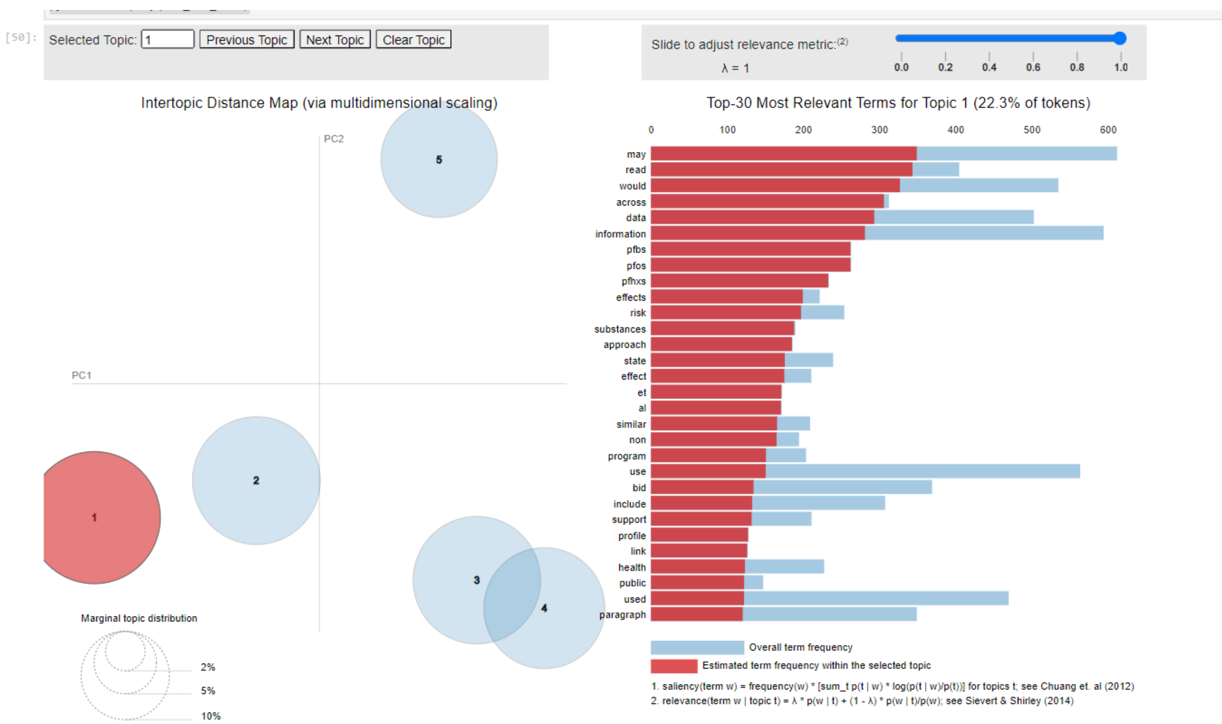

Word Cloud for Topic 4




   - The distribution of LDA topics across the dataset was also visualized, helping us understand the prevalence of each topic.




   - We interpreted each topic by analyzing the key terms and summarizing the likely discussions within that topic.




**Output of topic summaries**

Topic 1:

- Key Terms: mtg, mw, meetings, committees, book, list, access, need, ii, wireless
- Interpretation: Likely covers aspects related to: mtg, mw, meetings...

Topic 2:

- Key Terms: section, also, update, page, general, delete, table, indexes, contents, statement
- Interpretation: Likely covers aspects related to: section, also, update...

Topic 3:

- Key Terms: table, page, project, line, bonds, requirements, must, column, subsection, one
- Interpretation: Likely covers aspects related to: table, page, project...

Topic 4:

- Key Terms: study, work, research, also, section, existing, end, subjects, information, completed
- Interpretation: Likely covers aspects related to: study, work, research...

Topic 5:

- Key Terms: may, read, would, across, data, information, pfos, pfbs, pfhxs, effects

- Interpretation: Likely covers aspects related to: may, read, would...

**Our Interpretation**

**Topic 1:**

**Key Terms:** "mtg, MW, meetings, committees, book, list, access, need, ii, wireless."

**Interpretation:** We understand this topic focuses on meetings and related logistics. It likely discusses scheduling, organizing committees, or the need for access to resources like books or wireless connections. It might be centered around the coordination of various tasks or groups.

**Topic 2:**

**Key Terms:** "section, also, update, page, general, delete, table, indexes, contents, statement."

**Interpretation:** This topic concerns the structure and organization of documents or reports. It could involve updates to specific sections, managing tables of contents, or general editorial tasks like deleting outdated information or updating indexes.

**Topic 3:**

**Key Terms:** "table, page, project, line, bonds, requirements, must, column, subsection, one."

**Interpretation:** This topic revolves around project documentation, emphasizing tables, specific requirements, and various subsections within a document. It might be discussing the detailed aspects of a project, including necessary bonds and columns of data.

**Topic 4:**

**Key Terms:** "study, work, research, section, existing, end, subjects, information, completed."

**Interpretation:** I gather that this topic concerns research studies or projects. It likely discusses the progress of these studies, their completion, and any related documentation sections. The focus could be on summarizing or finalizing research work.

**Topic 5:**

**Key Terms:** "may, read, would, across, data, information, photos, pubs, pfhxs, effects."

**Interpretation:** This topic is about the analysis and interpretation of data. It might involve discussing the effects of certain substances (like PFOS or PFBS) and disseminating this information. The topic could be focusing on the implications of this data across various contexts.

## Results

### Zero-Shot Learning:

- The model achieved an accuracy of approximately 47.33%, with a precision of 49.67%, recall of 47.33%, and F1 score of 42.09%. It frequently labeled comments as "important," which may indicate a tendency to generalize across different topics.

```
Accuracy: 0.47325185333600484
Precision: 0.49669816947247025
Recall: 0.47325185333600484
F1 Score: 0.4208836468269039
```

### N-Shot Learning:

- After fine-tuning, the DistilBERT model achieved a validation accuracy of 69.07%, with precision, recall, and F1 scores indicating a balanced performance. This demonstrates the model's ability to improve with even a limited amount of labeled data.

| Epoch | Training Loss | Validation Loss | Accuracy | F1 | Precision | Recall |
|---|---|---|---|---|---|---|
| 1 | No log | 1.320747 | 0.515516 | 0.364303 | 0.433476 | 0.515516 |
| 2 | No log | 1.071949 | 0.595596 | 0.510645 | 0.577812 | 0.595596 |
| 3 | No log | 0.820963 | 0.690691 | 0.633566 | 0.595691 | 0.690691 |

### LDA Topic Modeling:

**Topic Identification:** The LDA model successfully identified critical topics within the comments, ranging from general operational themes to specific research-related discussions.

**Comparison with Zero-Shot and N-Shot Labels:** The comparison between LDA topics and the labels generated by zero-shot and N-shot models revealed that while LDA provided a good overview of the topics, the labeling methods sometimes needed to align perfectly with these topics. This suggests that further refinement of the labeling approaches could enhance the accuracy of automated classifications.

## Conclusion and Next Steps

### Insights Gained:

- The zero-shot and N-shot models provided a valuable baseline for automated comment classification, but the results highlighted areas for improvement, particularly in label alignment with underlying topics.

- The LDA topic modeling offered a deeper understanding of the content, revealing the need for more nuanced label generation.

**Future Work:**

- The next phase will develop a triage system that integrates the insights from zero-shot, N-shot, and LDA analyses. This system will prioritize and categorize comments more effectively, enabling more accurate and meaningful classifications.

- We will also work on refining our models to align the generated labels with the actual content, improving the overall accuracy and utility of the system.

This phase successfully explored and demonstrated the potential of advanced machine-learning techniques for comment classification. The findings and insights will be instrumental in developing a robust triage system in the next phase, further enhancing the automation of collaborative document management.