Q1.

```
ROW 1
{'tank': 0.47, 'plant': 0.8875, 'perplace': 0.5575, 'smsspam': 0.3471223021582734}

ROW 2
{'tank': 0.805, 'plant': 0.89, 'perplace': 0.8475, 'smsspam': 0.3471223021582734}

ROW 3
{'tank': 0.805, 'plant': 0.89, 'perplace': 0.8475, 'smsspam': 0.3471223021582734}

ROW 4
{'tank': 0.805, 'plant': 0.89, 'perplace': 0.8475, 'smsspam': 0.3471223021582734}

ROW 5
{'tank': 0.76, 'plant': 0.91, 'perplace': 0.85, 'smsspam': 0.3471223021582734}

ROW 6
{'tank': 0.805, 'plant': 0.875, 'perplace': 0.8475, 'smsspam': 0.3471223021582734}

ROW 7
{'tank': 0.865, 'plant': 0.91, 'perplace': 0.6525, 'smsspam': 0.9676258992805755}

ROW 8
{'tank': 0.8975, 'plant': 0.9025, 'perplace': 0.85, 'smsspam': 0.9676258992805755}

ROW 9
{'tank': 0.8975, 'plant': 0.9025, 'perplace': 0.85, 'smsspam': 0.9676258992805755}

ROW 10
{'tank': 0.8975, 'plant': 0.9025, 'perplace': 0.85, 'smsspam': 0.9676258992805755}

ROW 11
{'tank': 0.9025, 'plant': 0.9225, 'perplace': 0.85, 'smsspam': 0.9676258992805755}

ROW 12
{'tank': 0.89, 'plant': 0.87, 'perplace': 0.8375, 'smsspam': 0.9676258992805755}
```

Q2.
Here's a breakdown of the provided text in simpler terms:

1. Weighting Approach:
   - The method used is a modified version of stepped weighting.
   - It extends the consideration to a wider range of words and adjusts their weights accordingly.
   - Different weights are assigned based on the distance of words from the target term.

2. Rationale for Weighting Values:
   - The weights are determined based on the belief that words closer to the target have more importance.

- Once the context spans 10 words (5 on each side of the target), further words have less impact.
- The weights were adjusted to find the best values, leading to varied results with each change.
- Further testing is needed on larger datasets to optimize the weighting values.

3. Best Model Findings:
- The preferred model consists of standard term frequency, overlap similarity, and no removal of stopwords.
- Notable improvement was seen in the smsspam task when using overlap similarity instead of cosine similarity, significantly boosting accuracy to around 95%.
- Stopwords removal slightly improved some tasks but worsened smsspam and perplace tasks significantly, thus were retained for a holistic approach.
- Stemming was found to consistently worsen performance compared to a model without stemming.

Q3.

1. Targeted task:
- In a targeted task, like identifying an ambiguous word or a named entity, a specific target term exists.
- The target term is handled separately and does not carry weight in the model.
- It serves as a focal point for refining the model, enabling improvements such as scaling surrounding word weights and extracting contextual information and special adjacent tokens.
- These enhancements are expected to significantly boost task performance.

2. Non-targeted Task :
- In contrast, non-targeted tasks treat all terms equally without differentiation.
- This approach lacks the ability to establish context or apply positional weighting.
- Without a specific target term to guide adjustments, the model may struggle to discern the importance of individual words or their positions in the text.

3. Importance of Target Term:
- The presence of a target term allows for more nuanced adjustments and improvements in the model.
- By focusing on the target term, the model can better understand the surrounding context and assign appropriate weights to words.
- This targeted approach is particularly beneficial for tasks where specific elements need to be identified or analyzed within a larger body of text.

4. Enhanced Model Performance:
- Utilizing a target term often leads to enhanced model performance by providing a clear focal point for analysis and refinement.

- By incorporating the target term into the model's decision-making process, it can more accurately interpret the text and produce more reliable results.
- This targeted approach may be especially valuable in tasks requiring precise identification or classification of key elements within textual data.

Q4.

1. Doubts arose regarding term weighting in the smsspam task due to the absence of a specific target word like 'plant', making position weighting challenging. Consequently, a uniform weight assignment of 1 to every term was chosen, disregarding position or context.

2. Reflection revealed potential flaws in the chosen approach. For example, initial words in a text message may hold more importance, influencing immediate viewer engagement as they're typically displayed in text notifications.

3. Conversely, words at the end of the text, especially those containing common spam keywords, might carry more weight. Prioritizing such keywords through heavier weighting could enhance spam message identification.

4. Acknowledging these considerations suggests adjustments to the weighting scheme could improve spam detection accuracy. Considering word importance at both text beginning and end, along with specific keywords, may enhance the model's effectiveness.

5. Text messages exhibit distinct characteristics where initial words set the message tone, while concluding words provide crucial context or action items. Reflecting these nuances in weighting strategies can enhance spam identification in text messages.

6. Spam messages evolve with new tactics and keywords to evade detection. Adapting the weighting scheme to prioritize frequently occurring spam keywords or patterns can help the model keep pace with emerging spam trends.

7. Utilizing machine learning techniques like feature engineering or neural networks allows the model to learn and adapt its weighting scheme based on training data. This adaptive approach improves the model's understanding of spam indicators, potentially enhancing performance over time.

8. Incorporating user feedback can provide insights into words or phrases users perceive as indicative of spam. This iterative process allows continuous refinement of the weighting scheme, leading to more accurate spam detection based on real-world usage and user preferences.