**Key for interpreting Canary Smoking Output “TobaccoUse V002.6”**

**Classification is made at the sentence level and then the note level.** **Each line of output represents one sentence.** A frequency approach, followed by a hierarchy approach if there’s a frequency tie, for both sentence-level and overall note-level classification should be taken.

**Steps:**

1. Delete all “1’s” from lines that are followed in the very next line/sentence by a 7, 10, 13, 17, 21, or 22.
2. Delete all “1’s” that are in the same sentence with a 7 or 19.
   1. Here are examples of how to apply steps 1 and 2:

|  |  |  |
| --- | --- | --- |
| **Sentence** | **Canary Output** | **How to treat the “1”s.** |
| Line 1 | 1,15 | Consider |
| Line 2 | 1 | Ignore / delete because followed by a 7 in the very next sentence. |
| Line 3 | 7,1 | Ignore / delete because in the same sentence as 7. |

1. Examine each sentence and assign classification to the **sentence** based on the output table here:

|  |  |
| --- | --- |
| 1, if **not** followed in the next sentence by 7, 10, 13, 17, 21, or 22.  If 1 is followed in the next sentence by 7, 10, 13, 17, 21, or 22, the 1 should be ignored. | Current |
| 2 | Current |
| 3 | Current |
| 4 | Current |
| 5-11, 13-14, 20-21 | Past |
| 12 | Past? Haven’t seen output for this. |
| 15-19, 22-23 | Non-smoker |
| 1,19 in same sentence (order of output does not matter) | Non-smoker |
| 1, 7 in same sentence (order of output does not matter) | Past |

* 1. If a sentence has multiple types of output, use a frequency and then hierarchy approach to assign smoking type at the sentence level.
     1. Choose the smoking status that is encountered most frequently for that sentence.
        1. For example, if there is *one* output of *current* smoking and *two* outputs indicating *past* smoking, the sentence will be counted as that of a *past* smoker.
     2. If there is a tie between frequencies, a hierarchy of current, past, non-smoker should be applied.
        1. For example, if a sentence has *one* mention of *current* smoking and *one* mention of *past* smoking, the sentence will be classified as *current* smoking.

1. Once all sentences have been assigned smoking status, apply the same frequency and then hierarchy approach at the **note level.**
   1. Count the output (1 category from each sentence) associated with each type of smoking status per note.
   2. Choose the smoking status that is encountered most frequently for that note.
      1. For example, if there is *one* count of *current* smoking and *two* sentences indicating *past* smoking, the note will be counted as that of a *past* smoker.
   3. If there is a tie between frequencies, a hierarchy of current, past, non-smoker should be applied.
      1. For example, if a note has *one* mention of *current* smoking and *one* mention of *past* smoking, the note will be classified as *current* smoking.