Table 1 Part-of-Speech tags for verbs

Tag	Definition
VB	base
VIP	present tense, not 3rd person singular
VBZ	Present tense, 3rd person singular
MED	Past tense
VUG	Present Participle
VEN	past participle

filter out such words with the help of a POS tagger, 2) A POS tagger can also be used to distinguish words that can be used in different parts of speech. For instance, as a verb, "enhanced" may conduct different amount of sentiment as being of an adjective. The POS tagger used for this research is a max-entropy POS tagger developed for the Penn Treebank Project [31]. The tagger is able to provide 46 different tags indicating that it can identify more detailed syntactic roles than only 8. As an example. Table 1 is a list of all tags for verbs that has been included in the POS tagger.

Each sentence was then tagged using the POS tagger. Given the enormous amount of sentences, a Python program that is able to run in parallel was written in order to improve the speed of tagging. As a result, there are over 25 million adjectives, over 22 million adverbs, and over 56 million verbs tagged out of all the sentiment sentences, because adjectives, adverbs, and verbs are words that mainly convey sentiment.

Negation phrases identification

Words such as adjectives and verbs are able to convey opposite sentiment with the help of negative prefixes. For instance, consider the following sentence that was found in an electronic device's review: "The built in speaker also has its uses but so far nothing revolutionary. The word, "revolutionary" is a positive word according to the list in [27]

Algorithm 1 Negation Phrases Identification

Require: Tagged Sentences, Negative Prefixes

Ensure: NOA Phrases, NOV Phrases

1. **for** every Tagged Sentences **do**
2. **for** i/i + 1 as every word/tag pair **do**
3. **if** i+ 1 is a Negative Prefix **then**
4. **if** there is an adjective tag or a verb tag in next pair **then**
5. NOA Phrases ← $(i, i+2)$
6. NOV Phrases ← $(i, i+2)$
7. **else**
8. **if** there is an adjective tag or a verb tag in the pair after next **then**
9. NOA Phrases ← $(i, i+2, i+4)$
10. NOV Phrases ←(i, i + 2, i + 4)
11. **end if**
12. **end if**
13. **end if**
14. **end for**
15. **end for**
16. **return** NOA Phrases, NOV Phrases

Table 2 Top 10 sentiment phrases based on occurrence

Phrase	Type	Occurrence
not worth	NOA	26329
not go wrong	NOA	15446
not bad	NOA	15112
not be happier	NOA	14892
not good	NOA	12919
don't like	NOV	42525
didn't work	NOV	38287
didn't like	NOV	21806
don't work	NOV	10671
don't remember	NOV	9670

$$ss(t) = \frac{\sum_{i=1}^{5} i \times \gamma 5,i \times Occurrenc \quad i(t)}{\sum_{i=1}^{5} \gamma 5,i \times Occurrenc \quad i(t)} \tag{2}$$

Occurrence$_i$(t) is *t*'s number of occurrence in *i*-star reviews, where *i*=1,….,5. According to Figure 3, our dataset is not balanced indicating that different number of reviews were collected for each star level. Since 5-star reviews take a majority amount through the entire dataset, we hereby introduce a ratio,$\gamma 5, i$,which is defined as:

$$\gamma 5, i = \frac{|5-star|}{|i-star|} \tag{3}$$

In equation 3, the numerator is the number of 5-star reviews and the denominator is the number of *i*-star reviews, where *i* = 1,…..,5. Therefore, if the dataset were balanced, $\gamma 5, i$ would be set of 1 foe every *i*. Consequently, every sentiment score should fall into the interval of [1,5]. For positive word tokens, we expect that the median of their statement scores should exceed 3, which is the point of being neutral according to figure 1. For negative word tokens, it is to expect that the median should be less then 3.