|  |  |  |  |
| --- | --- | --- | --- |
| **Scenario** | **Entropy** | **Purity (Performance)** | **Note** |
| 1. Base Model | 0.646 | 0.425 | Unigrams only |
| 2. Base Model + (Intersecting vocabulary  with Google's pre-trained model) | 0.425 | 0.614 | Unigrams only |
| 3. Base Model + (Conceptualizing text and  Intersecting vocabulary  with Conceptualizer's pre-trained model) | 0.354 | 0.681 | ngrams (1-8) |
| 4. Base Model + (Conceptualizing text and  Intersecting vocabulary  with Conceptualizer's pre-trained model) + (Document shuffling for Patent2Vec training) | 0.352 | 0.697 | ngrams (1-8) |