Appendix A: Full Experiment Results

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Step 1** | | | | | | | | | |
| **id** | **topic model** | **process** | **filtered ?** | **tech stop words** | **num topics** | **cv** | **npmi** | **umass** | **uci** |
| 0 | lda | lemma | TRUE | FALSE | 9 | 0.378 | 0.007 | -2.033 | -0.278 |
| 1 | lda | lemma | TRUE | TRUE | 9 | 0.451 | 0.008 | -3.546 | -0.782 |
| 2 | lda | stem | TRUE | FALSE | 9 | 0.381 | 0.011 | -2.126 | -0.293 |
| 3 | lda | stem | TRUE | TRUE | 9 | 0.462 | 0.025 | -2.514 | -0.186 |
| 4 | nmf | lemma | TRUE | FALSE | 9 | 0.378 | 0.014 | -1.740 | 0.052 |
| 5 | nmf | lemma | TRUE | TRUE | 9 | 0.554 | 0.062 | -2.618 | 0.113 |
| 6 | nmf | stem | TRUE | FALSE | 9 | 0.420 | 0.024 | -1.770 | 0.033 |
| 7 | nmf | stem | TRUE | TRUE | 9 | 0.529 | 0.062 | -2.066 | 0.394 |
| 8 | lda | lemma | FALSE | FALSE | 9 | 0.440 | 0.016 | -2.385 | -0.310 |
| 9 | lda | lemma | FALSE | TRUE | 9 | 0.461 | 0.001 | -3.639 | -0.989 |
| 10 | lda | stem | FALSE | FALSE | 9 | 0.450 | 0.026 | -2.003 | -0.033 |
| 11 | lda | stem | FALSE | TRUE | 9 | 0.446 | -0.001 | -3.325 | -0.843 |
| 12 | nmf | lemma | FALSE | FALSE | 9 | 0.389 | 0.001 | -2.066 | -0.458 |
| 13 | nmf | lemma | FALSE | TRUE | 9 | 0.534 | 0.037 | -2.838 | -0.588 |
| 14 | nmf | stem | FALSE | FALSE | 9 | 0.417 | 0.015 | -1.891 | -0.211 |
| 15 | nmf | stem | FALSE | TRUE | 9 | 0.510 | 0.043 | -2.388 | -0.134 |
|  |  |  |  |  |  |  |  |  |  |
| **Step 2** | | | | | | | | | |
| **id** | **topic model** | **phrasing** | **filtered dataset** |  | **num topics** | **cv** | **npmi** | **umass** | **uci** |
| 0 | lda | FALSE | TRUE |  | 9 | 0.479 | 0.037 | -3.284 | -0.255 |
| 1 | lda | TRUE | TRUE |  | 9 | 0.527 | -0.174 | -3.580 | -5.431 |
| 2 | lda | FALSE | FALSE |  | 9 | 0.481 | 0.025 | -3.726 | -0.636 |
| 3 | lda | TRUE | FALSE |  | 9 | 0.587 | -0.185 | -5.947 | -5.865 |
| 4 | nmf | FALSE | TRUE |  | 9 | 0.578 | 0.072 | -3.032 | 0.046 |
| 5 | nmf | TRUE | TRUE |  | 9 | 0.565 | -0.143 | -2.657 | -4.570 |
| 6 | nmf | FALSE | FALSE |  | 9 | 0.679 | 0.113 | -2.421 | 0.705 |
| 7 | nmf | TRUE | FALSE |  | 9 | 0.689 | -0.134 | -2.493 | -4.589 |
|  |  |  |  |  |  |  |  |  |  |
| **Step 3** | | | | | | | | | |
| **id** | **topic model** | **n-gram** | **filtered dataset** |  | **num topics** | **cv** | **npmi** | **umass** | **uci** |
| 1 | lda | 1 | TRUE |  | 9 | 0.464 | 0.038 | -1.826 | 0.219 |
| 2 | lda | 2 | TRUE |  | 9 | 0.416 | 0.023 | -1.914 | 0.053 |
| 3 | lda | 3 | TRUE |  | 9 | 0.402 | 0.009 | -2.377 | -0.249 |
| 4 | lda | 1 | FALSE |  | 9 | 0.494 | 0.047 | -1.851 | 0.316 |
| 5 | lda | 2 | FALSE |  | 9 | 0.459 | 0.037 | -1.936 | 0.234 |
| 6 | lda | 3 | FALSE |  | 9 | 0.427 | 0.029 | -1.855 | 0.170 |
| 7 | nmf | 1 | TRUE |  | 9 | 0.464 | 0.037 | -1.892 | 0.173 |
| 8 | nmf | 2 | TRUE |  | 9 | 0.408 | 0.021 | -1.960 | 0.074 |
| 9 | nmf | 3 | TRUE |  | 9 | 0.414 | 0.023 | -1.951 | 0.084 |
| 10 | nmf | 1 | FALSE |  | 9 | 0.479 | 0.040 | -1.940 | 0.171 |
| 11 | nmf | 2 | FALSE |  | 9 | 0.422 | 0.023 | -2.035 | 0.003 |
| 12 | nmf | 3 | FALSE |  | 9 | 0.418 | 0.026 | -1.962 | 0.093 |

Appendix B: List of Identified Technical Stop Words

|  |  |  |  |
| --- | --- | --- | --- |
| process | accuracy | analysis | present |
| level | compare | one | technique |
| develop | detection | feature | data |
| measure | perform | order | two |
| network | new | learn | study |
| increase | apply | however | development |
| conduct | require | service | determine |
| many | provide | performance | high |
| power | work | result | also |
| various | different | improve | type |
| need | make | signal | frequency |
| number | simulation | approach | method |
| propose | problem | research | several |
| indonesia | methods | user | design |
| applications | experiment | implement | test |
| use | system | show | obtain |
| algorithm | technology | time | information |
| reduce | - | find | model |
| build | quality | communication | value |
| software | implementation | measurement | support |
| base | condition | image | control |
| systems | paper | area | application |