**Lexical Analysis of Agatha Christie Texts**

|  |  |  |
| --- | --- | --- |
| **Lexical Features** | **Before 1945**  **Mean(SD)** | **After 1945**  **Mean(SD)** |
| Unique word | 20.55(24.34) | 22.9(27.13) |
| Noun | 25.04(14.74) | 26.47(17.49) |
| Pronoun | 6.99(2.04) | 6.8(2.5) |
| Conjunction | 3.01(0.86) | 3.1(1.01) |
| Indefinite Term | 1.15(0.35) | 1.2(0.47) |
| Reduced Sentences | 4.03(0.81) | 3.8(1.2) |
| Interjections | 0.20(0.11) | 0.22(0.11) |
| Total Word Count | 50210.70(24612.01) | 48556.11(28402.02) |
| Adjectives | 7.17(1.29) | 7.04(1.58) |
| Verb | 19.65(4.23) | 20.00(3.03) |
| Number of sentences | 5089.68(2513.85) | 4744.639(2837.638) |
| Adverb | 6.82(1.89) | 6.91(2.18) |
| TTR | 0.20(0.24) | 0.22(0.3) |
| Brunet’s Index | 12.90(1.87) | 12.82(2.3) |
| HONORE’s statistic | 997.92(236.85) | 976.01(256.37) |
| Mean word length | 594.51(841.70) | 796.902(1250.76) |
| Lexical Density | 58.70(9.63) | 60.44(12.19) |
| Function words | 21.82(5.46) | 20.56(6.46) |
| Punctuation | 8987.78(4355.03) | 8457.83(5060.79) |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| **Lexical Features** | **Before 1960**  **Mean(SD)** | **After 1960**  **Mean(SD)** |
| Unique word | 23.12(27.16) | 17.49(20.26) |
| Noun | 26.29(16.35) | 24.04(15.20) |
| Pronoun | 6.76(2.34) | 7.38(1.97) |
| Conjunction | 3.02(0.97) | 3.16(0.82) |
| Indefinite Term | 1.12(0.40) | 1.29(0.41) |
| Reduced Sentences | 3.95(1.04) | 3.86(0.95) |
| Interjections | 0.20(0.11) | 0.24(0.10) |
| Total Word Count | 48156.684(26138.764) | 53086.4(27043.04) |
| Adjectives | 7.34(1.27) | 6.45(1.68) |
| Verb | 19.52(4.05) | 20.65(2.31) |
| Number of sentences | 4861.754(2650.82) | 5118.2(2737.72) |
| Adverb | 6.73(2.09) | 7.25(1.80) |
| TTR | 0.23(0.27) | 0.17(0.20) |
| Brunet’s Index | 12.72(2.08) | 13.28(2.05) |
| HONORE’s statistic | 976.06(256.87) | 1020.79(208.89) |
| Mean word length | 720.19(1135.18) | 600.62(778.06) |
| Lexical Density | 59.90(11.25) | 58.40(9.87) |
| Function words | 21.27(6.26) | 21.11(5.07) |
| Punctuation | 8581.91(4638.16) | 9190.6(4868.015) |
|  |  |  |

**Results from model evaluation**

We considered the works before 1960(denoted by class no) as to be written in normal mind and that after 1960(denoted by class yes) to be written in diseased mind. **10 fold cross validation.**

|  |  |  |
| --- | --- | --- |
| **Model** | **Correctly classified Instances** | **Mean absolute Error** |
| Bayes Net | 94.8 | 0.07 |
| Random Committee | 96.10 | 0.12 |
| REP tree | 97.40 | 0.03 |
|  |  |  |
|  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **FP Rate**  **(yes/no)** | **TP Rate**  **(yes/no)** | **Precision**  **(yes/No)** | **Recall**  **(yes/No)** | **F (yes/No)** | **PRC Area (yes/No)** | **ROC Area (yes/No)** |
| Bayes  Net | 0.035/ 0.10 | 0.90/ 0.96 | 0.90/  0.96 | 0.90/ 0.96 | 0.90/ 0.965 | 0.92/ 0.94 | 0.93/ 0.93 |
| Random Committee | 0.01/ 0.10 | 0.90/ 0.98 | 0.94/ 0.96 | 0.90/ 0.98 | 0.92/ 0.97 | 0.94/ 0.97 | 0.96/ 0.96 |
| REP Tree | 0.01/ 0.05 | 0.95/ 0.98 | 0.95/ 0.98 | 0.95/ 0.98 | 0.95/ 0.98 | 0.93/ 0.97 | 0.96/ 0.96 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

**Features Selection (Using Information Gain)**

1. Verb
2. Indefinite terms
3. Nouns
4. Interjections
5. Reduced Sentences
6. Pronoun
7. Conjunctions
8. Adjectives
9. Lexical Density
10. Honore’s statistics
11. Brunnett’s index
12. Adverb
13. TTR
14. Unique words

