Rule Count: 70

Rule Miss: 11 (15%)

Rule Statistics

|  |  |  |  |
| --- | --- | --- | --- |
|  | Rules | Miss (Mario dataset) | Miss (Ruby dataset) |
| CA | 36 | 14 (39%) | 14 (39%) |
| CD | 25 | 18 (72%) | 3 (12%) |
| D | 8 | 6 (75%) | 1 (13%) |
| CH | 1 | 0 (100%) | 0 (100%) |
| Total | 70 | 38 (54%) | 18 (25%) |

The rules were created in the Ruby dataset and tested in both Ruby and Mario dataset. The tables shows that the rules created in Ruby dataset is not effective in Mario dataset as 54% of the rules missed, while 25% of the rules missed in the Ruby dataset. For the CA category, the rules with the highest was the <pos:PSNS><ner:LOCATION>[as]LOCATION, with 3530 total hits. This rule extracts the format “ : <location”. The hits came from the Ruby dataset with 3282 hits. The rules missed 39% for both dataset. This is primarily from the retweets instances of the Ruby dataset. For the CD category, the rules with the highest hits is the <ner:LOCATION>[as]LOCATION <pos:PSNS>, with 111 total hits. The rules mostly missed on the CD category, with 72% missed rate. For the Donation category, only 8 rules were created. This is because there were only few instances collected for this category. Lastly for the CH rules, only 1 rule was created because of too few instances and no relevant information could be collected.

Rules with the highest hits per category

|  |  |  |
| --- | --- | --- |
| Category | Rule | Hits |
| CA | <pos:PSNS><ner:LOCATION>[as]LOCATION | 3530 |
| CD | <ner:LOCATION>[as]LOCATION <pos:PSNS> | 111 |
| D | <pos:IN><ner:LOCATION>[as]LOCATION | 24 |