|  |  |  |  |
| --- | --- | --- | --- |
| S.no | No. of  URL’s | Features Left  (after 1st Reduction) | Time  (after normalizing )  sec.ms |
| 1 | 50 | 1061 | 1.149 |
| 2 | 100 | 1765 | 1.517 |
| 3 | 500 | 5398 | 3.593 |
| 4 | 1000 | 9384 | 7.112 |
| 5 | 2000 | 14909 | 3.227 |
| 6 | 5000 | 28266 | 30.435 |
| 7 | 20000 | 74113 | Not completed |

Time taken to enqueue 20 URL out of 5000 = 0.053ms and for dequeue 10 URL = 0.007ms. The top 40 URL all belong to class +1.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| UID1 | IMP1 | UID2 | IMP2 | UID3 | IMP3 | UID4 | IMP4 | UID5 | IMP5 | UID6 | IMP6 | UID7 | IMP7 |
| 6 | 7.827 | 38 | 7.732 | 33 | 7.618 | 12 | 7.362 | 5 | 4.932 | 19 | 4.544 | 35 | 4.143 |
| 6 | 4.349 | 38 | 4.317 | 12 | 4.172 | 33 | 4.062 | 85 | 3.968 | 83 | 3.841 | 74 | 3.747 |
| 6 | 4.609 | 291 | 4.578 | 38 | 4.385 | 33 | 4.342 | 288 | 4.342 | 134 | 4.342 | 213 | 4.274 |
| 784 | 3.432 | 689 | 3.432 | 123 | 3.230 | 98 | 3.221 | 382 | 3.115 | 296 | 3.107 | 421 | 3.107 |
| 28 | 3.242 | 1020 | 3.217 | 1114 | 3.217 | 335 | 3.122 | 1031 | 3.094 | 609 | 3.080 | 1233 | 3.003 |
| 2309 | 3.482 | 2214 | 3.482 | 2799 | 3.456 | 3257 | 3.441 | 2903 | 3.435 | 3014 | 3.428 | 3015 | 3.428 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| S.no | Features taken after finding IG values of features. | No. of URL’s | UID1 | UID2 | UID3 | UID4 | UID5 | UID6 | UID7 | UID8 | UID9 | UID10 | Time  mm:ss |
| 1 | 100 | 100 | 28 | 23 | 5 | 27 | 83 | 26 | 21 | 57 | 90 | 13 | 02:01 |
| 2 | 200 | 100 | 28 | 5 | 83 | 23 | 27 | 26 | 21 | 57 | 13 | 90 | 02:02 |
| 3 | 300 | 100 | 28 | 5 | 83 | 23 | 27 | 26 | 21 | 57 | 13 | 90 | 01:59 |
| 4 | 500 | 100 | 73 | 83 | 50 | 49 | 84 | 47 | 85 | 43 | 42 | 41 | 01:58 |
| 5 | 1000 | 100 | 73 | 83 | 50 | 49 | 84 | 47 | 85 | 43 | 42 | 41 | 02:10 |

In the 3rd table it was observed that when features taken out after finding IG values(2nd column) was 500 or greater than 500, the importance value of top 10 features was same, for 200 and 300 the result was same for top 10 features, so we have set our threshold =200, i.e. after finding IG values we are taking out 200 features.

Sir make this correction everywhere that we are taking top 200 features and not top 100 .

2nd table is the continuation of first table, use dekh ke adjust kar lena, aur ye bhi mention kar dena ki the first two tables are for day0 and 3rd is for day1