Table **Error! No text of specified style in document.**‑. Results of Mario No-Retweet dataset (Using 10 Cross-Fold Validation)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F-Measure | Kappa |
| Random Forest | 0.947 | 0.945 | 0.942 | 0.8429 |
| J48 | 0.908 | 0.911 | 0.899 | 0.7365 |
| KNN3 | 0.896 | 0.899 | 0.886 | 0.6894 |
| KNN5 | 0.833 | 0.867 | 0.835 | 0.5498 |
| KNN7 | 0.826 | 0.847 | 0.809 | 0.4496 |
| Naïve Bayes | 0.926 | 0.928 | 0.926 | 0.8063 |
| Bayesian Network | 0.852 | 0.893 | 0.87 | 0.6789 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F-Measure | Kappa |
| Random Forest | 0.947 | 0.945 | 0.942 | 0.8429 |
| J48 | 0.908 | 0.911 | 0.899 | 0.7365 |
| KNN3 | 0.896 | 0.899 | 0.886 | 0.6894 |
| KNN5 | 0.833 | 0.867 | 0.835 | 0.5498 |
| KNN7 | 0.826 | 0.847 | 0.809 | 0.4496 |
| Naïve Bayes | 0.926 | 0.928 | 0.926 | 0.8063 |
| Bayesian Network | 0.852 | 0.893 | 0.87 | 0.6789 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F-Measure | Kappa |
| Random Forest | 0.947 | 0.945 | 0.942 | 0.8429 |
| J48 | 0.908 | 0.911 | 0.899 | 0.7365 |
| KNN3 | 0.896 | 0.899 | 0.886 | 0.6894 |
| KNN5 | 0.833 | 0.867 | 0.835 | 0.5498 |
| KNN7 | 0.826 | 0.847 | 0.809 | 0.4496 |
| Naïve Bayes | 0.926 | 0.928 | 0.926 | 0.8063 |
| Bayesian Network | 0.852 | 0.893 | 0.87 | 0.6789 |

Table **Error! No text of specified style in document.**‑. Results of Mario No-Retweet dataset (Using Test Split)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F-Measure | Kappa |
| Random Forest | 0.945 | 0.942 | 0.933 | 0.8322 |
| J48 | 0.913 | 0.906 | 0.89 | 0.7226 |
| KNN3 | 0.862 | 0.856 | 0.826 | 0.4906 |
| KNN5 | 0.825 | 0.842 | 0.8 | 0.4064 |
| KNN7 | 0.835 | 0.856 | 0.821 | 0.4786 |
| Naïve Bayes | 0.909 | 0.914 | 0.908 | 0.7654 |
| Bayesian Network | 0.923 | 0.921 | 0.907 | 0.7599 |

Table **Error! No text of specified style in document.**‑. Results of Ruby No-Retweet dataset (Using 10 Cross-Fold Validation)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F-Measure | Kappa |
| Random Forest | 0.884 | 0.892 | 0.882 | 0.7204 |
| J48 | 0.845 | 0.857 | 0.846 | 0.6381 |
| KNN3 | 0.825 | 0.826 | 0.794 | 0.4793 |
| KNN5 | 0.814 | 0.797 | 0.747 | 0.3453 |
| KNN7 | 0.782 | 0.775 | 0.709 | 0.2352 |
| Naïve Bayes | 0.858 | 0.827 | 0.832 | 0.6167 |
| Bayesian Network | 0.864 | 0.852 | 0.856 | 0.6676 |

Table **Error! No text of specified style in document.**‑. Results of Ruby No-Retweet dataset (Using Test Split)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F-Measure | Kappa |
| Random Forest | 0.892 | 0.892 | 0.882 | 0.7142 |
| J48 | 0.822 | 0.836 | 0.821 | 0.5609 |
| KNN3 | 0.781 | 0.722 | 0.795 | 0.2786 |
| KNN5 | 0.779 | 0.757 | 0.675 | 0.1317 |
| KNN7 | 0.757 | 0.749 | 0.658 | 0.0887 |
| Naïve Bayes | 0.868 | 0.852 | 0.859 | 0.6661 |
| Bayesian Network | 0.824 | 0.781 | 0.795 | 0.5232 |

Table **Error! No text of specified style in document.**‑1 and Table **Error! No text of specified style in document.**‑2 shows the result of the classification testing of the Mario no-retweets dataset. It shows that the classifier could classify unique tweets. In comparison with the 10 cross-fold validation and test-split, there is a little difference in the result. Random Forest could still correctly classify almost all of the instances. However, the other classifier could not classify the other category, especially the small instances. The other classifier could not classify the D category. However, there is a problem with the Ruby datasets, shown in Table **Error! No text of specified style in document.**‑3 and Table **Error! No text of specified style in document.**‑4. The dataset introduces the problem of unbalanced dataset. Although, the f-measure is high. The problem is now in the kappa statistics. Most of the classifier (KNN3, KNN5, KNN7) is below 0.5. They tend to classify all the instances to the O category.

Table **Error! No text of specified style in document.**‑. Datasets with No Retweet Statistics

|  |  |  |
| --- | --- | --- |
|  | Mario | Ruby |
| CA | 268 (77%) | 154 (16%) |
| CD | 19 (5%) | 74 (7.8%) |
| CH | 44 (12%) | 6 (0.6%) |
| D | 16 (4%) | 17 (1.7%) |
| O | NA | 696 (73.5%) |
| Total | 347 | 947 |

Table **Error! No text of specified style in document.**‑5 shows each content of the dataset with removed retweets. The table shows that the dataset is unbalanced, leaning to CA category for the Mario dataset and O category for the Ruby dataset. Also, in the Mario dataset, there are no O category because all of the instances that were gathered is a retweet.