For <table>, this uses the top 10% features for each category. The table shows that the Random Forest got the highest score with 0.84 F-measure and 0.7201 kappa statistics, while Naïve Bayes with 0.719 F-measure and 0.5141 kappa statistics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | Precision | Recall | F-measure | Kappa |
| J48 | 0.812 | 0.813 | 0.807 | 0.6636 |
| Random Forest | 0.841 | 0.842 | 0.84 | 0.7201 |
| kNN-3 | 0.792 | 0.791 | 0.786 | 0.6245 |
| kNN-5 | 0.781 | 0.779 | 0.779 | 0.6001 |
| kNN-7 | 0.77 | 0.768 | 0.759 | 0.5771 |
| Naïve Bayes | 0.731 | 0.725 | 0.719 | 0.5141 |
| Bayesian Network | 0.752 | 0.752 | 0.747 | 0.5553 |

For <table>, this uses the top 20% features for each category. The table shows that Random Forest has the highest F-measure and kappa statistics, scoring 0.854 and 0.942 respectively, while Naïve Bayes scored the lowest with 0.726 F-measure and 0.5372 kappa statistics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | Precision | Recall | F-measure | Kappa |
| J48 | 0.82 | 0.819 | 0.814 | 0.675 |
| Random Forest | 0.855 | 0.856 | 0.854 | 0.942 |
| kNN-3 | 0.796 | 0.794 | 0.789 | 0.6293 |
| kNN-5 | 0.784 | 0.782 | 0.776 | 0.606 |
| kNN-7 | 0.78 | 0.78 | 0.773 | 0.6015 |
| Naïve Bayes | 0.751 | 0.73 | 0.726 | 0.5372 |
| Bayesian Network | 0.751 | 0.73 | 0.726 | 0.862 |

For <table>, this uses the top 30% features for each category. The table shows that Random Forest has the highest F-measure and kappa statistics, scoring 0.86 and 0.7527 respectively, while Naïve Bayes scored the lowest with 0.734 F-measure and 0.5464 kappa statistics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | Precision | Recall | F-measure | Kappa |
| J48 | 0.833 | 0.834 | 0.831 | 0.7043 |
| Random Forest | 0.86 | 0.862 | 0.757 | 0.7527 |
| kNN-3 | 0.798 | 0.796 | 0.788 | 0.6293 |
| kNN-5 | 0.794 | 0.793 | 0.786 | 0.6244 |
| kNN-7 | 0.807 | 0.804 | 0.798 | 0.6451 |
| Naïve Bayes | 0.751 | 0.731 | 0.734 | 0.5464 |
| Bayesian Network | 0.739 | 0.733 | 0.73 | 0.5365 |

Based on the results, the random forest algorithm is consistently the highest among the classifiers in all three settings, while Naïve Bayes is consistently the lowest. All the algorithm with the exception of Bayesian Network increase in precision and recall as the number of features increase. The reason that the Bayesian Network is suffering in precision and recall is because of the increasing network. The increasing number of attributes introduces a lot of noise.