**Classification Algorithm Analysis Table**

The table below show the results received when testing the algorithms used and their different scores in terms of Precision, Recall and F-Measure.

Please note:

* Oversampling has been applied on the dataset for this analysis.
* We are using binary feature to represent the presence of the word in a sentence
* Results were produced from both written algorithm and WEKA

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm | Training Set or Cross Validation (10 Folds) | Positive Class (risky) | | | Negative Class (non-risky) | | | Result Buffer from WEKA |
| Precision | Recall | F-Measure | Precision | Recall | F-Measure |
| Random Forest | Training Set | 1.000 | 0.994 | 0.997 | 0.998 | 1.000 | 0.999 | <ORFTS1> |
| Random Forest | Cross Validation | 0.980 | 0.897 | 0.937 | 0.969 | 0.994 | 0.982 | <ORFCV1> |
| Neural Network | Training Set |  |  |  |  |  |  | Produced in written algorithm |
| Neural Network | Cross Validation |  |  |  |  |  |  |
| Naïve Bayes | Training Set | 0.500 | 0.704 | 0.585 | 0.896 | 0.782 | 0.835 | <ONBTS1> |
| Naïve Bayes | Cross Validation | 0.481 | 0.688 | 0.566 | 0.889 | 0.770 | 0.825 | <ONBCV1> |

**No Oversampling Applied**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm | Training Set or Cross Validation (10 Folds) | Positive Class (risky) | | | Negative Class (non-risky) | | | Result Buffer from WEKA |
| Precision | Recall | F-Measure | Precision | Recall | F-Measure |
| Random Forest | Training Set | 1.000 | 0.987 | 0.993 | 0.999 | 1.000 | 0.999 | <RFTS1> |
| Random Forest | Cross Validation | 0.780 | 0.150 | 0.251 | 0.938 | 0.997 | 0.967 | <RFCV1> |
| Decision Trees | Training Set | 0.890 | 0.554 | 0.683 | 0.967 | 0.995 | 0.980 | <DTTS1> |
| Decision Trees | Cross Validation | 0.519 | 0.319 | 0.395 | 0.949 | 0.977 | 0.963 | <DTCV1> |
| Naïve Bayes | Training Set | 0.236 | 0.840 | 0.368 | 0.985 | 0.790 | 0.877 | <NBTS1> |
| Naïve Bayes | Cross Validation | 0.231 | 0.827 | 0.362 | 0.983 | 0.788 | 0.875 | <NBCV1> |
| Support Vector Machines | Training Set | 0.939 | 0.648 | 0.767 | 0.973 | 0.997 | 0.985 | <SVMTS1> |
| Support Vector Machines | Cross Validation | 0.556 | 0.420 | 0.479 | 0.956 | 0.974 | 0.965 | <SVMCV1> |