

## Справка по алгоритмам

1. Relative\_majority – самый наивный: подсчитываем для каждого элемента из test set, сколько аналогичных ему (то есть совпадающих по всем признакам) элементов присутствует в плюс ( $n^+$ ) и минус ( $n^-$ ) контекстах. Далее подсчитываем относительные величины  $n^+/N_+$  и  $n^-/N_-$ , где N – количество элементов в контексте. Далее относим элемент к тому классу, в котором вышеуказанная относительная величина больше. Столь «жесткая» классификация не позволяет классифицировать ни один объект и мы получаем {'not\_classified': 92, 'contradictory': 0, 'positive\_negative': 0, 'negative\_negative': 0, 'positive\_positive': 0, 'negative\_positive': 0}

Логичным продолжением такого вида классификации является возможность разрешения неточного соответствия, то есть для каждого классифицируемого объекта (x) вычисляется количество объектов из плюс и минус контекста, отличающихся от x не более, чем на z признаков. z подбирается вручную, с целью максимизировать долю верно распознанных объектов. Целевой показатель – accuracy – доля верно распознанных объектов. Перебрав все возможные значения z, выяснилось, что при z = 1 многие объекты не классифицируются, максимальный показатель accuracy (98.9%) достигается при z = 2, а следующий при z = 8 (89.01%)

```
{'negative_positive': 0.0, 'not_classified': 0.0, 'negative_negative': 31.999999999999996,
'contradictory': 0.0, 'positive_negative': 0.9999999999999999, 'positive_positive':
57.999999999999996}
```

Sensitivity (TPR) of the procedure is 98.31%

Fall-out (FPR) of the procedure is 0.0%

Precision (PPV) of the procedure is 100.0%

False omission rate (FOR) of the procedure is 3.03%

Accuracy of the procedure is 98.9%

2. Алгоритм, основанный на расстоянии Хэмминга: два объекта удалены друг от друга на столько, каково для них количество несовпадающих признаков. После вычисления среднего расстояния Хэмминга для классифицируемого объекта он относится к тому классу, расстояние Хэмминга до которого меньше.

```
{'contradictory': 0.0, 'not_classified': 0.0, 'negative_negative': 19.0, 'negative_positive':
13.000000000000002, 'positive_negative': 23.000000000000004, 'positive_positive':
36.000000000000001}
```

Sensitivity (TPR) of the procedure is 61.02%

Fall-out (FPR) of the procedure is 40.63%

Precision (PPV) of the procedure is 73.47%

False omission rate (FOR) of the procedure is 54.76%

Accuracy of the procedure is 60.44%