

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) посчитывается 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 – ассурасу – доля верно распознанных объектов. Перебрав все возможные значения z , выяснилось, что при $z = 1$ многие объекты не классифицируются, максимальный показатель ассурасу (98.9%) достигается при $z = 2$, а следующий при $z = 8$ (89.01%)
- 2.