1. По сравнению с результатами для английского языка, получилось хуже.

Результаты для классификации по 4 фичам:

**RandomForest=49,5%**

J48=48,5%

DecisionTable=44%

SimpleLogistic=42,5%

SMO=39,5%

NaiveBayes=37%

One-R=37%

По сравнению с английским языком получилось хуже. Возможно причина в том, что в крпусе всего 200 вхождений; и в некоторых (7-8) примерах не удалось указать значение, так как в корпус попадали названия и имена собственные (например, партия «Жизнь»), в следствие чего приходилось указывать нулевое значение.

1. На мой взгляд, датасет недостаточно сбалансирован. Из-за этого в выдаче мы можем наблюдать явное преимущество 4 значения.
2. Лучше всего:4 ,хуже всего 1,7 и 8.
3. Улучшилось.

Результаты классификации с ключевыми словами:

**DecisionTable=95,5%**

**SimpleLogistic=95,5%**

One-R=95%

SMO=95%

RandomForest=93,5%

J48=92%

NaiveBayes=92%

По всем показателям качество значительно улучшилось после добавления ключевых слов. Однако лидирующие позиции заняли DecisionTable, SimpleLogistic