Значения слова «бой»:

1.Битва, сражение. ---------*Наступательные бои. Воздушный бой. Поле боя. Разведка боем. Вступить в бой*

2.Борьба, состязание. ----------*Кулачный бой. В больших боях участвовал*

3.Действие по глаг. бить, битье, побои --------------- *боем дело не исправишь*

4.Убой --------------*Бой рыбы*

5.Убойная способность огнестрельного оружия ------------ *Проверить бой автомата*

6.Битье предметов ---------- *Бой посуды*

7.Собир. – Разбитая, расколотая на куски посуда, предметы ------------- *Кирпичный бой*

8.Звуки этого действия ------------------ *Барабанный бой. Бой часов*

9.Бой – использование иностранного «boy» ------------ *Американ бой. Геймбой. Секси бой*

10. Имя собственное

Ключевые слова соответсвуют вышеупомянутым значениям (то есть под первым числом – ключевые слова, соответсвующие значению «битва, сражение»):

1.вступить, воздушный, ринуться, морской, нападение, отступление, смертный, смертельный

2.кулачный, рукопашный, соперник

3.

4.рыбы, животных, оленя

5.автомата, пулемета

6.посуда, тарелки

7.кирпичный

8.барабанный, куранты, часы

9.американ, спайс, гейм

10.

В «датасете с группами» классы “one”,”two”,”three” и т.д. соответствуют значениям в том же порядке, т.е. класс “one” соответствует значению «битва, сражение». Если в конкретном предложении встречается одно из вышеупомянутых ключевых слов, то в классе (“one”,”two”и т.д.) напротив это предложения будет стоять единица. Если же ключевых слов из той группы значений в предложении не найдено, то будет стоять 0. Поэтому в основном в датасете везде стоят нули, ведь ключевые слова не так часто встречаются.

Ответы на задание:

1.Качество получилось ужасное. Связано это с тем, что в словаре MAC у двух самых популярных значений «битва, сражение»(1) и «борьба,состязание»(2) почти идентичные слова-соседи: предлоги «в» и «рваться» перед словом и предлог «с» после. Более того, четкой границы между этими двумя значениями найти не получилось: непонятно, когда «Клава рвется в бой» она представляет себе кулачную борьбу или идет, как на войну? С другой стороны складывать классы тоже не совсем корректно, так как первое значение подразумевает более жестокие обстоятельства, чем второе: все-таки война и состязание в виде спорта имеют мало общего между собой.

2.Датасет оказался несбалансирован: больше 70% значений были сфокусированы на первом и втором классе (и я не уверен, что я распределил значения между этими классами полностью правильно), значений 3,4,5 и 7 не было в выборке вовсе. Также пришлось добавить довольно спорные 9 и 10 пункты, отслеживанию которых обучить очень трудно (точнее легко, если просто перечислить все американизмы(9) и фамилии, связанные с именем Бой(10)). С другой стороны, среди такого количества 1-ых и 2-ых классов с хорошей частотой угадываются звуки (8) и битье предметов (6)

3. Хуже всего – второе значение, так как у него самая размытая формулировка (не считая тех значений, которые вообще не встретились). Лучше всего различается первый и восьмой классы

4.Лучше всего указывают ключевые слова 8 и 9 пунктов – там однозначные ответы, когда ключевое слово встречается в предложении. Также хорошо работают ключевые слова из первого класса