# 1. Выберите себе многозначное слово в [этой таблице](https://docs.google.com/spreadsheets/d/19VRKiVOvL4R_40tnag6LCbL5SOjvl59puH_ggMb0Wz4/edit?usp=sharing), запишите его за собой.

Свет

# 2. Составьте список значений слова, взяв в качестве основы толкования из словаря МАС

|  |  |  |  |
| --- | --- | --- | --- |
| **Значение** | **Примеры** | **Количество** | **Ключевые слова** |
| 1. **Электромагнитное излучение, воспринимаемое глазом и делающее видимым окружающий мир.** | *-Солнечный свет.*  *-Свет свечи.* | 66 | Солнечный, лунный, бледный, ослепительный, красный, зеленый, желтый, фонарь, свеча, яркий, электрический, луна, солнце |
| **2. Освещение, характерное для какой-л. части суток.** Рассвет, восход солнца. | *-Дневной свет.*  *-На улицах был еще неопределенный, рассеянный свет белых ночей.*  *-Еще семь часов до света!*  *-Как и в прошлую ночь, едва забрезжил свет, она уже проснулась.* | 8 | Чуть, ни, дневной, забрезжить |
| **3. Источник освещения и приспособление для освещения в домах и на улицах.** | *-Зажечь в комнате свет.* | 60 | Гореть, включать, включить, погаснуть, гаснуть, гасить, тушить, верхний, погасить, зажечь, зажечься, зажигаться |
| **4. Место, откуда исходит освещение, освещенное место, пространство, где светло.** | *-Встать спиной к свету.*  *-Доктор взял ребенка на руки, быстро повернул к свету и заглянул в глаза.* | 8 | на |
| **5. Светлое место, пятно на картине, передающее наибольшую освещенность какого-л. участка изображаемого пространства.** | *-Контрасты света и тени.* | 0 |  |
| **6. Блеск глаз под влиянием какого-л. чувства, радостное, ясное выражение лица.** | *-Все лицо его светилось тогда каким-то внутренним светом.* | 1 | Внутренний, исходить |
| **7. То, что делает ясным, понятным мир, то, что делает радостной, счастливой жизнь. Употребляется как символ истины, разума, просвещения или радости, счастья.** | *-Свет истины.*  *-Андрей стал говорить о свете знания, о святости труда, о широком и дружном товариществе.* | 15 | Истина, любовь, проливать, пролить, знание |
| **8. (*обычно со словом* «мой»*)* Ласковое, приветливое обращение к кому-л.** | *-Свет ты мой! послушай меня, старика: напиши этому разбойнику, что ты пошутил.* | 2 | Мой |
| 1. **Земля со всем существующим на ней, мир, вселенная.** | *-Части света.*  *-Путешествие вокруг света.*  *-И знаю, что лучше этого народа никакого народа на свете нет!*  *-Мигающий огонек горел на мысе Крильон, самом южном советском мысе в этой стороне света.* | 30 | На, новый, старый, увидеть, появляться, божий |
| 1. **Окружающие люди, общество. – не обязательное высшее** | *-И сердце бросил в море жизни шумной, И свет не пощадил — и бог не спас!*  *-Ане казалось, что весь свет видит ее дешевую шляпку и дырочки на ботинках.* | 14 | Белый, в |
| 1. **Ограниченный круг людей, составляющий высший слой привилегированных классов буржуазно-дворянского общества.** | *-Большой свет.*  *-Бывать в свете.*  *-Положение в свете.* | 5 | Высший |

# Вопросы

1. **Какого качества вам удалось достигнуть? Лучшее или худшее качество получилось в вашем случае по сравнению с результатами для английского языка?**

**RulesOneR – 66,17%**

* **Лучший результат показал** SMO – 66,6%

Результат работы других алгоритмов:

* SimpleLogistic – 66,17%
* RandomForest - 66,17%
* RandomForest – 66,17%
* MultiClassClassifier - 64,21%
* Logistic - 63,23%
* MultiLayerPerceptron – 62%
* RandomTree – 60,78%
* NaiveBayes – 59,31%

По сравнению с английским языком у меня получилось худшее качество. Думаю, что проблема связана с несбалансированностью выборки + кажется, что результаты были бы лучше, если бы у нас был корпус больше.

1. **Сбалансирован ли ваш датасет по количеству значений? Как это влияет на итоговый результат?**

Датасет не сбалансирован. Есть два очень больших класса, следующий по размеру класс в два раза меньше, остальные 7 совсем маленькие.

Из-за того, что датасет несбалансирован, классификатор часто ошибается в сторону двух самых больших классов – 1ого и 3-его, а классы, которые представлены чрезвычайно маленьким количеством примеров, классификатор зачастую вообще не распознает.

1. **Какие значения лучше всего различаются? Какие хуже?**

Лучше всего различаются значения, на которые было дано больше всего примеров – то есть 1ое и 3е. Хуже всего – 4ый, 8ой, 6ой классы – те, на которые было менее 10 примеров.

Алгоритм SMO лучше всего распознает 1 и 3 значения.

1. **Улучшилось ли качество для ключевых слов?**

Результаты по baseline не улучшились.

Baseline – 37.5%

Лучшая модель на прошлом датасете улучшилась:

RandomForest – 52,5%

Попробовала другие модели:

Tree: HoeffdingTree – 71.5%

Meta: IterativeClassifierOptimizer, MultiClassClassifierUpdateable – 72%

Lazy: LWL – 69.5%

Function: SMO – 73%

Bayes: NaiveBayes – 71.5%

Соответственно, для нового датасета лучшая модель SMO.

Может показаться, что я переусердствовала с ключевыми словами. Однако с движением может использоваться открытый класс слов, поэтому я выделила действительно ключевые слова, но они оказались частотны в моей выборке.