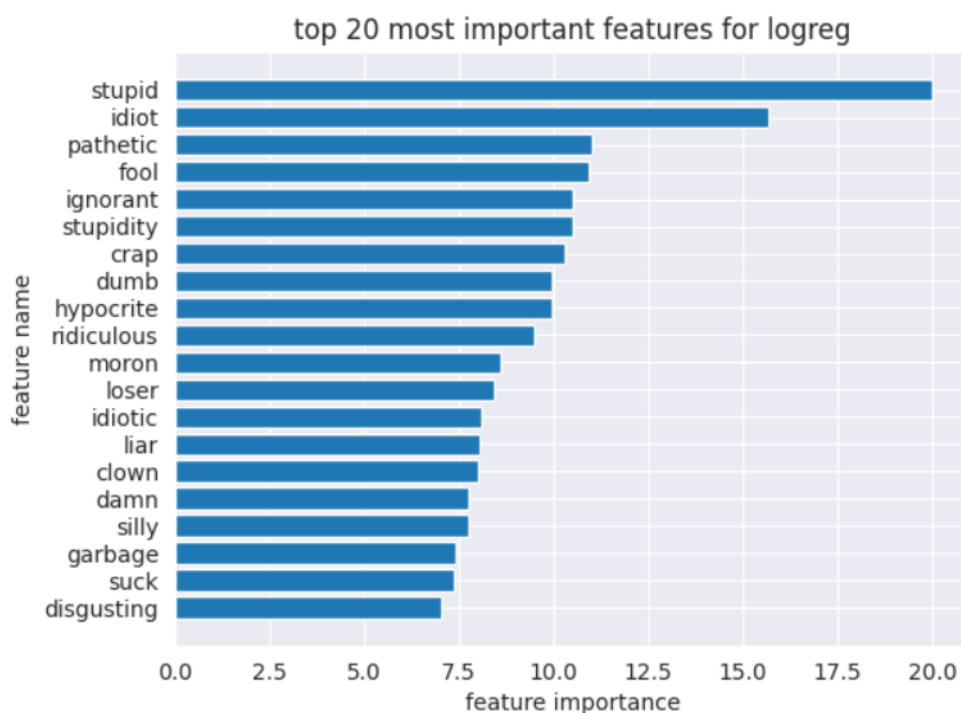


сравнение моделей:

модель	лучший скор после подбора гиперпараметров и определения порога	время предикта	порог
случайный лес	0.9235	6.32 s \pm 124 ms	0.4
логистическая регрессия	0.9317	2.12 ms \pm 69.6 μ s	0.4
дерево	0.8524	14.6 ms \pm 895 μ s	0.5
svm	0.8569	1min 32s \pm 399 ms	0.4
градиентный бустинг	0.8602	361 ms \pm 255 ms	0.4

Топ-20 наиболее важных при решении признаков довольно сильно пересекается среди моделей, а топ-5 совпадает (случайный лес, логрег и дерево) – это слова-индикаторы для токсичных сообщений.

В большей части алгоритмов наилучший результат был у моделей с порогом 0.4. На рисунке ниже представлены 20 наиболее существенных признаков в логистической регрессии с порогом 0.4:



После данного сравнения можно сделать вывод, что модель логистической регрессии наиболее предпочтительна – и по качеству, по времени предсказания.