

Самостоятельная работа номер 4. Вариант I

Решите все задания. Все ответы должны быть обоснованы. Все решения должны быть чётко приведены для каждого пункта. Рисунки должны быть чёткими, понятными. Все линии должны быть подписаны. Списывание карается обнулением баллов. Удачи!

[10] Задание 1

Всю свою жизнь Филипп был бездельником и сидел на шее у родителей. Но наконец ему улыбнулась удача и его взяли на работу аналитиком. Теперь он борется со всем плохим и за всё хорошее внутри большой рекомендательной системы.

Когда пользователь рекомендательной системы встречается в статьях матерную брань, он очень сильно расстраивается. Из-за этого одной из задач Филиппа является воспитание модели, которая могла бы дать бой нецензурщине.

Филипп оценил модель и получил на тестовой выборке следующие прогнозы вероятностей:

y_i	b_i
1	0.99
1	0.8
0	0.77
1	0.55
0	0.55
0	0.4
1	0.3
0	0.2

В первой колонке стоит 1, если статья была матерной и стоит 0, если статья оказалась обычной. Во второй колонке стоит вероятность того, что статья окажется матерной, предсказанная моделью. Филипп размышляет над тем какой бы для модели выбрать порог: 0.5, 0.7 или 0.9.

- (4) Найдите для каждого порога precision и recall. Какой порог вы бы порекомендовали поставить? Обоснуйте ответ.
- (4) Что такое roc-auc? Зачем изобрели эту метрику? Посчитайте её.
- (2) Что такое assigasy? Чем плоха эта метрика качества? Предложите конкретный численный пример, на котором эта метрика плохо себя показывает.