Отчет:

Атахужаев Собитхон

Программа состоит из следующих частей:

* Ищем оптимальный порог для улучшения точности нашей классификации и смотрим зависимость точности от нашего порога
* Рассматриваем зависимости насколько точность классификации зависит от объема наших тренировочных данных
* Берем исходные данные с сайта и сами создаем обучающую выборку и тестовую выборку чтобы распределить их равномерно(чтобы их количество были схожи)

**BHW\_Atakhujaev\_Sobitkhon.py**

* Наша программа в цикле изменяет значение порога и для каждого из значений порога запускает несколько раз программу которая считает количество правильно и неправильно отгаданных(правильные положительные, правильные отрицательные, неправильные положительные, неправильные отрицательные) классификаций. После этого из всех полученных точностей выбираем самый максимальный и для полученных результатов считаем Метрики качества.

Чтобы более понятно было каждая метрика рассмотрим простой пример и опишем каждую из метрик(Нам надо поймать терроситов).

|  |  |  |
| --- | --- | --- |
|  | TRUE | FALSE |
| TRUE | TP | FN |
| False | FP | TN |

TP-поймали террористов правильно

FN – количество пропущенных террористов (пропустили террориста)

FP – количество ложных тревог

TN- количество правильно не потревоженных мирных жителей

* True Positive Rate -  - “всех ли террористов поймали? Или процент правильно определенных террористов”
* True Negative Rate - “процент правильно определенных мирных жителей среди мирных”
* Negative Predictive Value “процент правильно пропущенных мирных жителей среди пропущенных”
* False Positive Rate-“процент потревожанных мирных ”
* False Discovery Rate-“процент мирных среди предсказанных террористов”
* Accuracy-“процент правильных ответов”
* Precision «сколько процентов террористов среди подозреваемых»
* Recall «»
* F1  «Средневзвешенное между precision и recall»

Самый максимальный результат был достигнут при пороге равной 0.9183884297520661.