|  |
| --- |
| Пензенский государственный университет  Факультет вычислительной техники  Кафедра «Вычислительная техника» |
| Отчет  по лабораторной работе №  по дисциплине «»  на тему «»  Вариант № 1 |
|  |
|  |
| Проверил:  Карамышева Н.С. |
| Пенза |

**Задание.**

1. На подготовленном наборе данных опробовать работу алгоритма ближайших соседей.

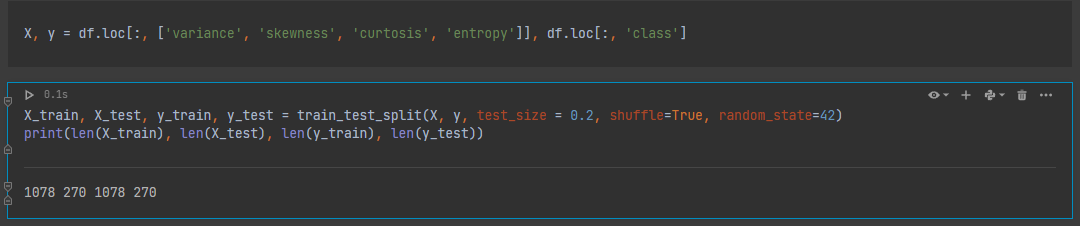
2. Провести оптимизацию гиперпараметров.

3. Описать использованные решения.

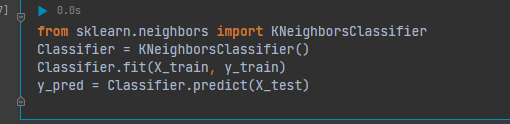
4. Оформить отчёт, содержащий информацию о выполнении пунктов 1-3 здания.

**Ход работы**

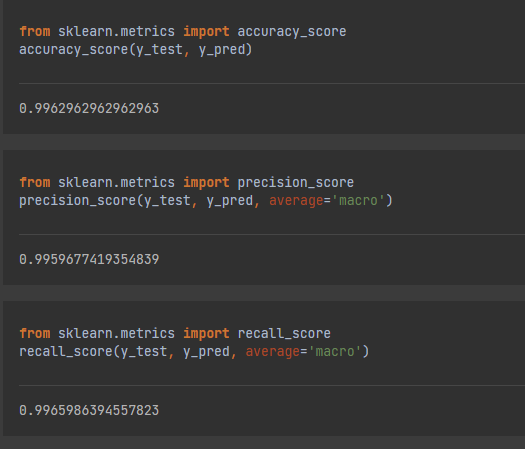
Разобьем данные на тренировочную и тестовую выборку в соотношение 80% на 20%.



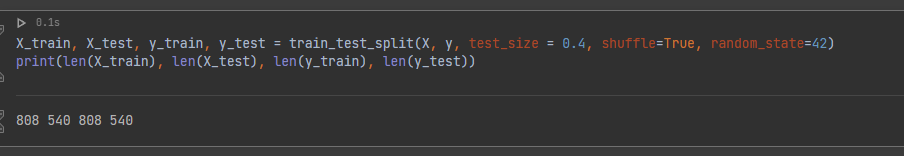
Воспользуемся методом ближайших соседей из библиотеки Sklearn.



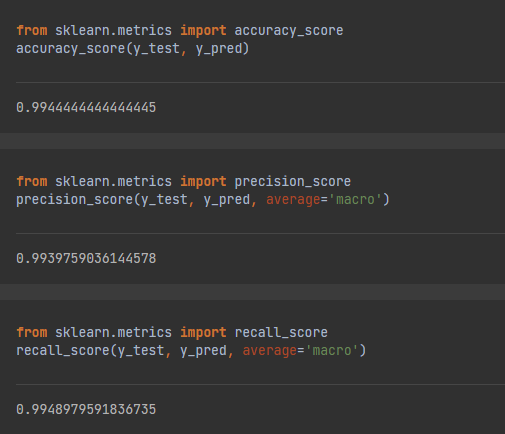
Теперь оценим работу классификатора с помощью трёх метрик: accuracy, precision и recall.



Отличный результат. Изменим отношение тренировочной и тестовой выборки на 60% к 40%.

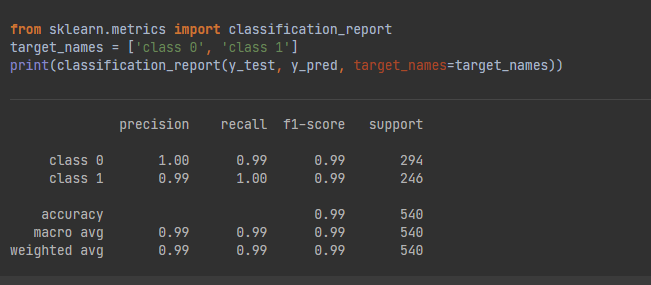


Результат метрик классификатора:



Результат также 99% при стандартных параметрах KNN (k = 5).

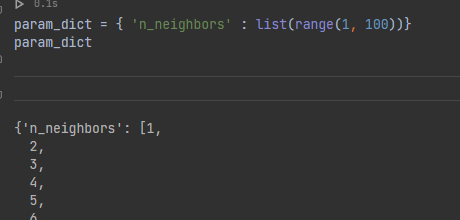
Также можно посмотреть classification\_report.



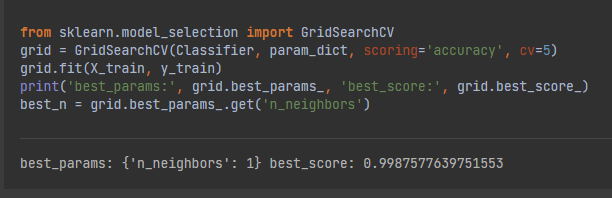
При таких высоких значениях метрик нет смысла подбирать параметры для классификатора, но по заданию это требуется.

Сделаем это с помощью GridSearchCV.

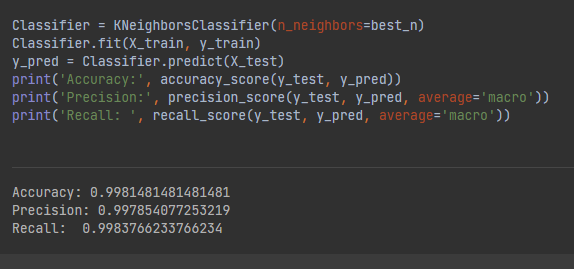
Создадим словарь с параметром n\_neightbors



И запустим GridSearchCV



По результату видно, что с параметром n\_neighbors = 1, наш классификатор должен выдать лучший результат.



Действительно, результат стал лучше, но на самом деле, оценка классификатора при подборке параметров проводилась на тренировочных данных, т.е на тех, которые использовались для обучения классификатора. Можно обернуть данный кусок кода в цикл и в нем ещё генерировать валидационную выборку, таким образом, результат станет ещё более независимым и правильным.

По результатам работы можно сказать, что полученные значения метрик превосходят ожидания и метод k ближайших соседей отлично справляется для решения подобной задачи.