Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)

Кафедра безопасности информационных систем (БИС)

Кейс «Аутентификация по параметрам динамики простановки подписи на графическом планшете»

Слушатель

Зотова Т.А.

12 декабря 2022

Принял канд.техн.наук,

доцент кафедры КИБЭВС

Е.Ю. Костюченко

12 декабря 2022

Томск 2022

Задание

При выполнении на сторонних средах (например, GoogleColab) необходимо выбрать файл sign\_dump\_part.csv, на локальной машине можно попробовать расширенный вариант с файлом sign\_dump.csv(если датасет на его основе поместится в памяти).

Т.е. проводится попарное комбинирование параметров и если записи в комбинации принадлежат одному пользователю, то выход 1, иначе 0.

Обращаем внимание на сильный дисбаланс в наборе – количество примеров, принадлежащих разным пользователям, больше практически в 10 раз.

Необходимо провести исследование минимум 3 различными классификаторами в рамках анализа набора данных для решения задачи классификации на 2 класса – 0 – половины параметров от разных пользователей и 1 – половины параметров от одного и того же пользователя.

Система оценивания – оценка проводится по средней F1-мере трех предложенных классификаторов.

**Набор данных**

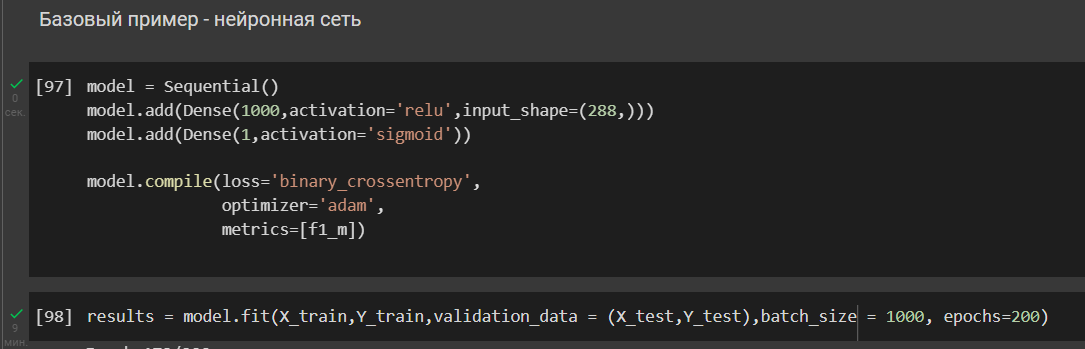
Размерность полученного набора данныхсоставляет 2045\*2045 запись из 144+144+1 полей, итого размер набора равен 4 182 025\*289.

Набор данных достаточно большой и в памяти помещается с трудом. Учитывая его большой размер и однородность по процедуре построения можно брать при построении классификатора не весь набор, а его часть (например, 100000 записей, количество можно варьировать в процессе исследования, но нужно не забыть их перемешать для случайного порядка). Итоговая оценка будет проводиться в 2 режимах: на 100000 записей, отбираемых случайным образом и, если позволяет память, полный набор. Полученное значение определяет получаемый балл.

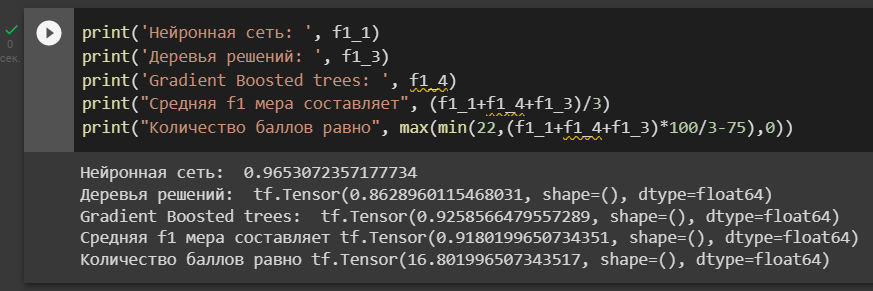
В файле 2\_Create\_Small.ipynb процедура обрезки набора с перемешиванием и выбором размера приведена, однако требует сформированного полного набора на предыдущем шаге.

**Ход работы и выводы**

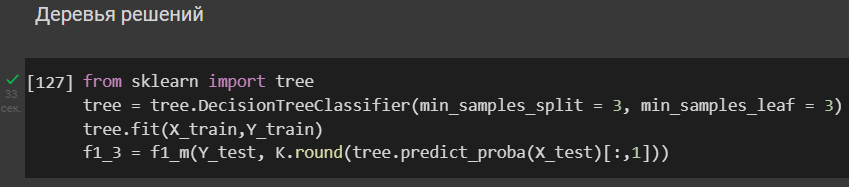
Построение классификатора на основе нейронной сети



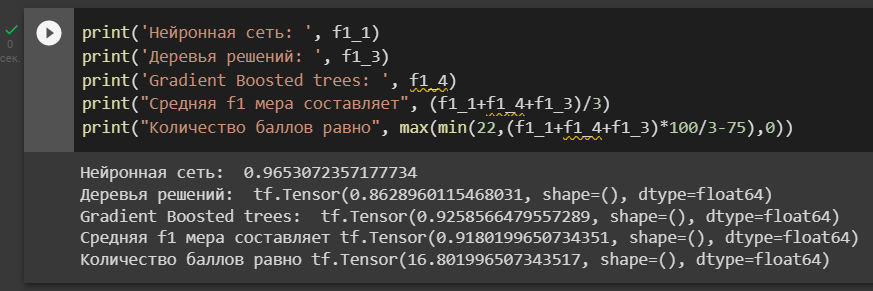
Подбор параметров на основе нейронной сети – значение F-меры - 0.9653



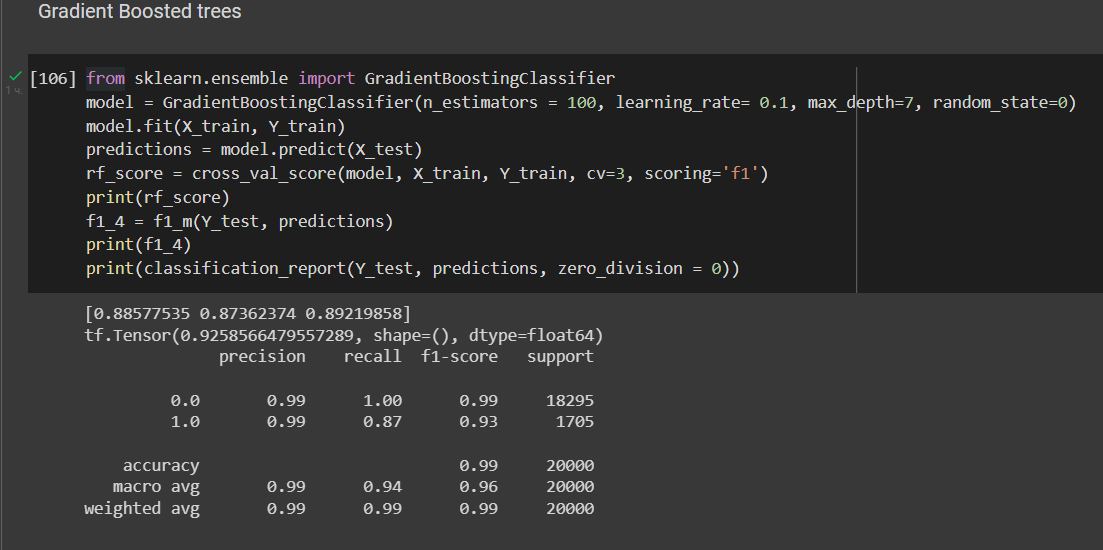
Построение классификатора на основе дерева решений



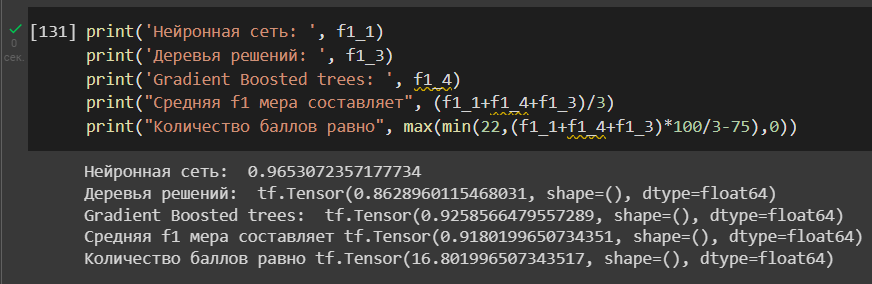
Подбор параметров на основе дерева решений – значение F-меры - 0.8628



Построение классификатора на основе GradientBoostedtrees



Подбор параметров на основе GradientBoostedtrees – значение F-меры - 0.9285



В результате проведенного эксперимента по подбору параметров видно, что исходя из перебранного множества значений можно рекомендовать выбор классификатора на основе нейронной сети, GradientBoostedtreesи Деревья решений, обеспечивающего следующие показатели качества работы системы: - Средняя f1 мера составляет 0.9180