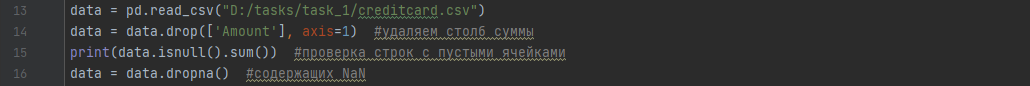
**Программирование в системах информационной безопасности  
 Лабораторный практикум**

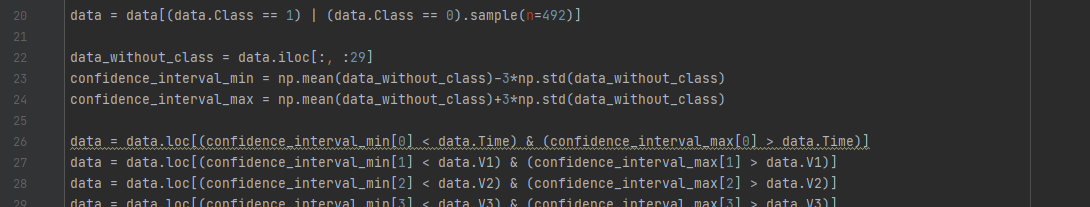
Васюткин Александр Олегович, группа 181-331

**Задание 3.**

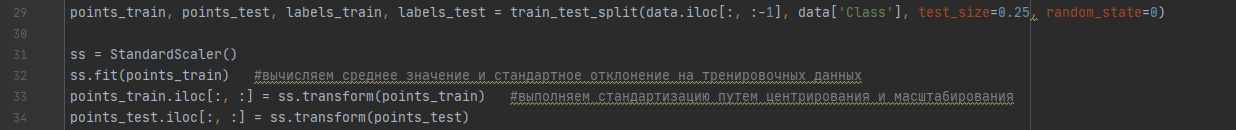
1. Предобработать данные.



После считывания файла программа выводит количество строк с пустыми ячейками и удаляет строки, содержащие NaN, а также столбец Amount, который был выявлен как малоинформативный в результате выполнения 7 лабораторной.

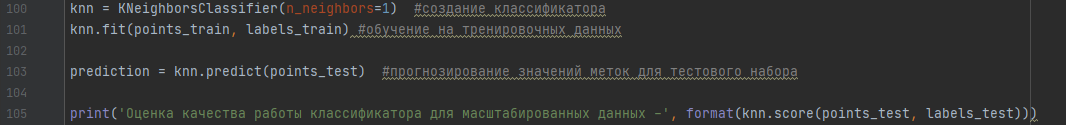


Далее в строке 20 уравниваем количество представителей различных классов, оставляя все записи с мошенническими транзакциями и такое же количество записей с не мошенническими. Затем ищем доверительный интервал и оставляем только те объекты, которые входят в него.



Делим признаки и метки на тренировочные и тестовые, затем масштабируем тренировочные и тестовые данные.

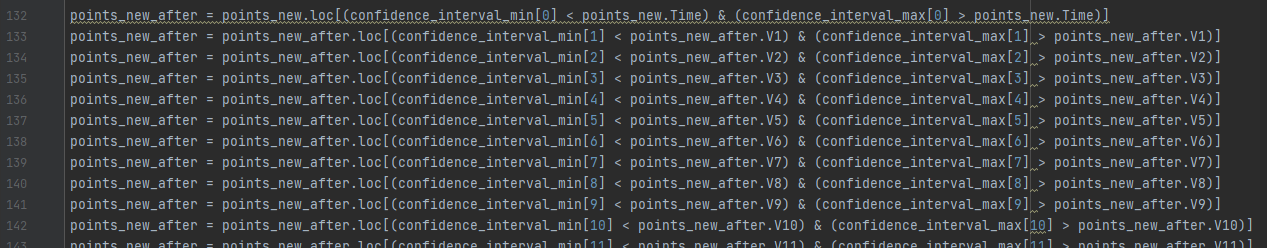
1. Построить и протестировать классификатор.



Создаем и обучаем классификатор на тренировочных данных. Протестировав классификатор на различном количестве соседей , я остановился на 1. Далее прогнозируем значения меток для тестового набора и выводим оценку качества работы классификатора.



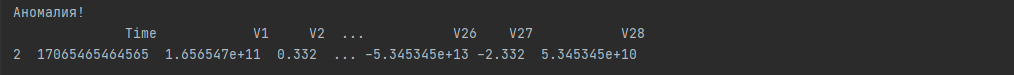
1. Реализовать метод, проверяющий значения признаков классифицируемого объекта на соответствие областям допустимых значений признаков и выявляющий аномальные объекты.



Просеиваем наши новые данные и записываем их в новый массив. Отбираем только те объекты, значения которых входят в доверительные интервалы, рассчитанные на этапе предобработки.



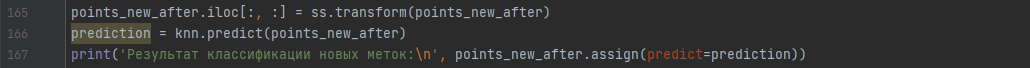
Далее, чтобы выявить аномальные объекты, мы находим строки, которые не прошли просеивание через доверительный интервал, путем нахождения различий между сетом со всеми исходными новыми данными и новыми данными, входящими в доверительный интервал.



1. Проиллюстрировать варианты эксплуатации классификатора.



Создаем сет из трех объектов, которые будем классифицировать. Первый объект я подобрал таким образом, чтобы он был похож на объект класса 0, второй – на объект класса 1, а для третьего данные были подобраны таким образом, чтобы он был определен как аномальный. Затем просеиваем эти данные через доверительный интервал (см. пункт 3).



Затем масштабируем новые данные, значения которых входят в доверительные интервал, классифицируем их и выводим результат классификации в консоль.

