МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение

высшего образования

«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра «Информационные системы»

Дисциплина «Методы искусственного интеллекта»

**Лабораторная работа № 3**

Построение и обучение нейронной сети для классификации

Выполнил:

студент гр. ИИмд-21

Шевалдов С.С

Проверил:

Доцент, к.т.н., доцент

Романов А.А.

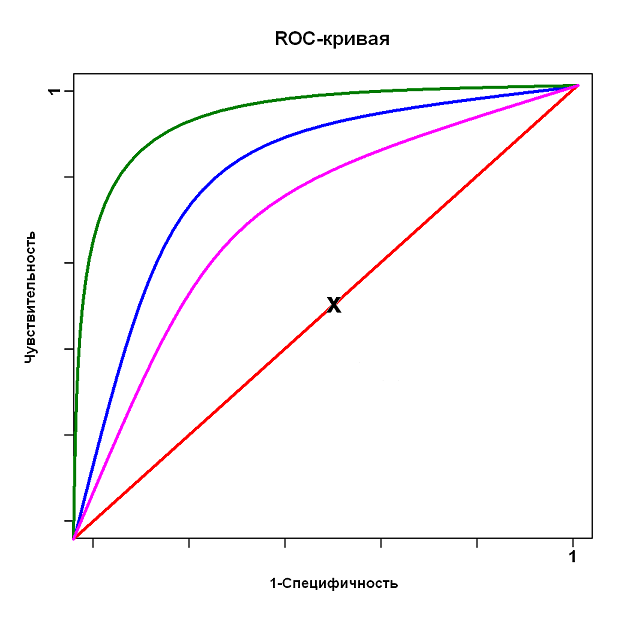
Ульяновск,   
2024 г.

**Задачи работы**

1. Ознакомиться с примером исходного кода на python, демонстрирующего определение фрода в банковских транзакциях.
2. Установить использованный метод для определения фрода. Ознакомиться с официальной документацией к нему.
3. Выбрать и загрузить один из сгенерированных наборов данных в среду google colab.
4. Выяснить установить влияние данных на меру качества ROC-AUC.
5. Составить отчет, описав зависимости от произведенных изменений.

**Порядок выполнения работы**

ROC-кривая описывает число верно классифицированных объектов по отношению к общему количеству объектов. Имеет значения в промежутке 0=<ROC=<1, где 0 – это все объекты были классифицированы ошибочно, и 1 - все объекты были классифицированы верно. Геометрический смысл – площадь под кривой в координатах от (0;0) до (1;1). Скажем в этом квадрате содержатся все объекты, которые надо классифицировать. И чем ближе кривая к диагонали, тем хуже результат работы классификации.



Зеленая линия захватывает большую площадь, следовательно качество классификации выше.

Зависимости от произведенных измерений:

|  |  |  |  |
| --- | --- | --- | --- |
| Имя файла | Количество строк | Количество столбцов | ROC-AUC |
| 4. trs1 | 10154 | 24 | 0,8625 |
| 7 | 0,4736 |
| 4. trs2 | 173832 | 24 | 0,7413 |
|  | 7 | 0,5658 |
| 4. trs3 | 16473 | 24 | 0,8283 |
|  | 7 | 0,409 |
| 4. trs5 | 32439 | 24 | 0,93 |
| 7 | 0,4798 |

В целом видно, что качество классификации слабо коррелирует с количеством строк в датасете. А вот количество столбцов в датасете напрямую влиет на качество классификации. Чем больше столбцов (информации о характеристиках транзакций), тем качественнее классификация.

В процессе обучения классификатора используется библиотека XGBClassifier. Она основана на принципах усиленного градиентного бустинга – обучении новой модели на основе ошибок предыдущих моделей для дальнейшего снижения ошибок.