## Лабораторная работа № 2

**Тема:** **«Классификация данных на основе методов классификации»**

**Цель работы:** разработка приложения для классификации объектов с использованием современных методов машинного обучения.

**Задачи:**

1. Загрузить данные с сайта: https://www.kaggle.com/paree24/development-index
2. Построить гистограммы с целью определения природы данных.
3. Задание в зависимости от варианта.
4. Визуализировать гистограммы прогнозируемых значений и реальных значений.
5. Отобразить f1-score и количество ошибок первого рода и второго рода.

**Варианты заданий и ссылки на библиотеки**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | Модель | Cross Validation | Ссылки на библиотеки и ресурсы | Ф.И.О студентов (группы) |
| 1 | Support Vector Machine | нет | https://scikit-learn.org/0.20/modules/svm.html#classification |  |
| 2 | Nearest Neighbors | да | https://scikit-learn.org/0.20/modules/neighbors.html |  |
| 3 | Random Forests | нет | https://scikit-learn.org/0.20/modules/ensemble.html#random-forests |  |
| 4 | AdaBoost | да | https://scikit-learn.org/0.20/modules/ensemble.html#adaboost | Бедердинова М.В.  Тюлюш Б.С.  Шамхалов Р.М. |
| 5 | Gradient Tree Boosting | да | https://scikit-learn.org/0.20/modules/ensemble.html#gradient-tree-boosting |  |
| 6 | Support Vector Machine | нет | https://scikit-learn.org/0.20/modules/svm.html#classification |  |
| 7 | Nearest Neighbors | нет | https://scikit-learn.org/0.20/modules/neighbors.html |  |
| 8 | Random Forests | да | https://scikit-learn.org/0.20/modules/ensemble.html#random-forests |  |
| 9 | AdaBoost | нет | https://scikit-learn.org/0.20/modules/ensemble.html#adaboost |  |

**Вопросы к лабораторной работе:**

* + - 1. Знание принципа работы выбранного алгоритма
      2. Дать определение ошибок 1 рода
      3. Дать определение ошибок 2 рода
      4. Дать определение точности и отзыва модели
      5. Дать определение f1 score