## Министерство науки и высшего образования РФ

***Hациональный исследовательский университет ИТМО***

## Факультет Программной инженерии и компьютерных технологий

**По дисциплине:**

Системы искусственного интеллекта

Лабораторная работа № 5.

*«Решение задачи многоклассовой*

*классификации - набор данных MNIST»*

**Выполнил**: До Зыонг Мань

**Группа**: Р33201

Санкт–Петербург 2022 год.

# Описание задания

Цель: решить задачу многоклассовой классификации, используя в качестве тренировочного набора данных - набор данных MNIST, содержащий образы рукописных цифр.

* 1. Используйте метод главных компонент для набора данных MNIST (train dataset объемом 60000). Определите, какое минимальное количество главных компонент необходимо использовать, чтобы доля объясненной дисперсии превышала 0.80 + номер\_в\_списке % 10. Построить график зависимости доли объясненной дисперсии от количества используемых ГК.
  2. Введите количество верно классифицированных объектов класса номер\_в\_списке % 9 для тестовых данных.
  3. Введите вероятность отнесения 5 любых изображений из тестового набора к назначенному классу.
  4. Определите Accuracy, Precision, Recall или F1 для обученной модели.
  5. Сделайте вывод про обученную модель. Вариант: Номер в списке 5.

Код можно посмотреть [здесь](https://colab.research.google.com/drive/1BNIcxvbKdEwJbwUAx85CGyls1HQepOmp#scrollTo=xyAb4Hi_FwrY)

# Выполнение

## Используйте метод главных компонент для набора данных MNIST

Минимальное количество главных компонент необходимо использовать, чтобы доля объясненной дисперсии превышала 0.85.

Text, application

Description automatically generated

График дисперсии:

Chart

Description automatically generated

## Введите количество верно классифицированных объектов

**Graphical user interface, text, application

Description automatically generated**

1. **Введите вероятность отнесения 5 любых изображений**

**Text

Description automatically generated**

Chart, histogram

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

Chart, histogram

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

A picture containing chart

Description automatically generated

## Определите Accuracy, Precision, Recall или F1

* 1. page2image31073424**Accuracy** - доля правильных ответов

Тем не менее, у этой метрики есть одна особенность, которую необходимо учитывать. Она присваивает всем документам одинаковый вес, что может быть не корректно в случае, если распределение документов в обучающей выборке сильно смещено в сторону какого-то одного или нескольких классов.

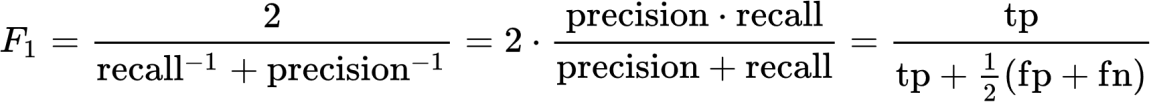
* 1. page2image31066768**Precision** – точность

Точность показывает, какая доля объектов, выделенных классификатором как положительные, действительно является положительными.

* 1. page2image31071760**Recall** – полнота

Полнота показывает, какая часть положительных объектов была выделена классификатором.

## F1 Score



Существует несколько способов получить один критерий качества на основе точности и полноты. Один из них — F-мера, гармоническое среднее точности и полноты

Table

Description automatically generated

# Вывод

Я использовал анализ основных компонентов, установил долю объясненной дисперсии выше 85% и обучил модель с основными компонентами 59. Общая точность этой модели составляет 80%, и модель распознает цифры 0, 1 точно высокая степень.