***Лабораторная работа № 3***

# **СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ КЛАССИФИКАЦИИ ДАННЫХ**

**Описание датасетов**

**1.** Набор данных ***«Iris Data Set»* – *«Ирисы Фишера»*** (см. Л.Р.№1)***.***

**Задание на лабораторную работу**

1. Провести снижение размерности методами МГК (PCA) и t-sne. Сделать предположение о разделимости выборки на классы. В отчёт включить все графики и вывод о разделимости.
2. Провести классификацию (с параметрами по умолчанию) (Классификаторы назначаются преподавателем). Проводить обучения классификатора следует с использованием кросс-валидации. Так же требуется разделить выборку на тестовую и обучающую. Так же требуется получить данные о качестве классификации (*Accuracy, Precision, Recall и F1-мера*). После обучения требуется визуализировать тестовую выборку:
3. С исходными метками
4. С метками полученными при классификации

В отчёт включить оба графика. Сделать выводы о качестве классификации.

1. Провести настройку параметров классификаторов с использованием gridsearchcv. Результаты настройки сохранить для следующего пункта. (При защите будет уделяться особое внимание параметрам классификации (обязательно знать, что каждый параметр обозначает) и принципу их обучения и работы)
2. Провести классификацию (с параметрами из предыдущего пункта). После обучения требуется визуализировать тестовую выборку:
3. С исходными метками
4. С метками полученными при классификации (пункт 2)
5. С метками полученными при классификации

В отчёт включить все графики. Сделать выводы о качестве классификации. Сравнить с результатами 2 пункта.

1. Исследовать возможность проведения классификации данных, содержащихся в датасетах: (Датасет назначается преподавателем)

- *Barotrop.sta*;

- *Chemical\_process.sta*;

- *Factor.sta*.