



## Специальные технологии баз данных и информационных систем

НИЯУ МИФИ, Кафедра финансового мониторинга.

Лабораторный практикум.

В.Ю. Радыгин, Д.Ю. Куприянов

### Семестр 2. Лабораторная работа 5

Лабораторная работа 5 рассчитана на два занятия и работу дома. Её целью является изучение основ классификации данных с помощью нейронных сетей в сравнение с классификаторами DecisionTree и SVM.

## Задание 1

1. Загрузите с сайта <https://sci2s.ugr.es/keel/datasets.php> набор статистических данных, указанный в вашем варианте. Разберитесь, какие данные приведены в наборе и какой атрибут является меткой класса.
2. На основе загруженного файла создайте Pandas DataFrame, подобрав правильные типы данных столбцов.
3. Выполните стандартизацию полученного дата фрейма.
4. Разделите дата фрейм на обучающую, тестовую и валидационную выборки в соотношении 5 / 3 / 2.
5. На основе обучающей и тестовой выборки постройте MLP-классификатор. Рассчитайте параметры его эффективности (Accuracy, Precision, Recall, ROC-AUC). Меняя значение параметров классификатора обоснованно подберите наиболее удачную MLP-сеть.
6. На основе обучающей и тестовой выборки постройте дерево решений, рассчитайте параметры эффективности классификатора (Accuracy, Precision, Recall, ROC-AUC). Меняя значение параметра альфа ([0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.2, 0.8]) и критерий классификации ([Entropy, Gini]) обоснованно подберите наиболее удачное дерево классификации для подготовленных выборок.
7. На основе обучающей и тестовой выборки постройте SVM-классификатор, рассчитайте параметры эффективности классификатора (Accuracy, Precision, Recall, ROC-AUC). Меняя значение параметров kernel, gamma, coef0, degree, C (на основе вариантов, представленных в лекции 1 второго семестра) обоснованно подберите наиболее удачное дерево классификации для подготовленных выборок.
8. Используя валидационную выборку рассчитайте для лучшего MLP-классификатора, дерева решений и лучшего SVM-классификатора параметры эффективности (Accuracy, Precision, Recall, ROC-AUC). Обоснуйте какой из двух классификаторов и когда лучше.

## Варианты

### Задание 1

1. <https://sci2s.ugr.es/keel/dataset.php?cod=210>
2. <https://sci2s.ugr.es/keel/dataset.php?cod=209>
3. <https://sci2s.ugr.es/keel/dataset.php?cod=107>
4. <https://sci2s.ugr.es/keel/dataset.php?cod=72>