**Лабораторная работа № 3**

**Тема**: Выделение признаков. Нормализация.

**Цель работы** *– Знакомство с методами предобработки данных.*

Для выполнения данной лабораторной работы можно использовать любой найденный датасет с числовыми значениями желательно разных порядков, **подходящий для задач классификации**. Можно использовать свой датасет из второй работы.

В примере к лабораторной использован датасет с Kaggle boston\_housing.csv. Он не подходит для задачи классификации, поэтому его не используйте.

1. Загрузите найденный датасет.
2. Очистите его от пропусков (если они есть) и переведите все признаки в числовой формат.
3. Сохраните обработанный датасет (например, под именем data1.csv)
4. Загрузите файл boston\_features.html.
5. Примените методы, описанные в этом файле к своему датасету.
6. Примените для выделения признаков алгоритм RandomForest (можно воспользоваться статьей по ссылке <https://www.analyticsvidhya.com/blog/2018/08/dimensionality-reduction-techniques-python/> )
7. Выполните нормализацию данных.
8. Сохраните обработанный датасет (например, под именем data2.csv)
9. Обучите любой метод классификации (DecisionTree или Knn . . . ) отдельно для data1.csv и для data2.csv (т.е. для датасетов до и после выделения признаков и нормализации). Рассчитайте точность модели. Сравните результаты.