

Инструкция по чтению онлайн-приложения ВКР на тему  
ФАКТОРЫ, ВЛИЯЮЩИЕ НА СТОИМОСТЬ ПЕРВИЧНОГО ЖИЛЬЯ:  
ПРЕДСКАЗАНИЕ ЦЕН ПРИ ПОМОЩИ ГЛУБОКОГО ОБУЧЕНИЯ

Алексеев А.Д., Жучков Д.А.

Онлайн-приложение содержит скрипты языка программирования Python, переведенные в PDF (изначально были в формате .py и .ipynb), а также данные в формате .xlsx, использовавшиеся на различных этапах работы. Файлы указаны в хронологическом порядке их применения.

1. Извлечение данных

- *JK\_urls\_with\_pages\_code.pdf* – список веб-страниц жилых комплексов с указанием количества страниц, с которых производился парсинг ссылок на объявления
- *get\_flat\_urls\_code.pdf* – код для извлечения ссылок на объявления квартир с сайтов жилых комплексов
- *flat\_urls.pdf* – список веб-страниц объявлений
- *flat\_parsing\_code.pdf* – код, при помощи которого производился парсинг информации с веб-страниц объявлений по их ссылкам
- *raw\_data.xlsx* – сырые данные, полученные в результате парсинга
- *get\_metro\_coordinates\_code.pdf* – код для извлечения координат станций метро из файла, скачанного с OpenStreetMap

2. Формирование переменных

- *calculate\_distance\_code.pdf* – код для расчета расстояния по прямой и по а/м дорогам

3. Предобработка

- *preprocessing\_code.pdf* – код, при помощи которого производилась чистка и преобразование данных
- *final\_data.xlsx* – итоговый датасет, на котором обучались модели

4. Построение моделей

- *choose\_nn\_city\_code.pdf* – код для «поиска по сетке» лучшей модели нейросети для всей выборки (Санкт-Петербург и Всеволожский район ЛО)
- *results\_nn\_city.xlsx* – результаты обучения моделей нейросети для всей выборки через перебор параметров
- *choose\_nn\_districts\_code.pdf* – код для «поиска по сетке» лучших моделей нейросети отдельно для каждого района
- *results\_nn\_districts.xlsx* – результаты обучения моделей нейросети отдельно для каждого района через перебор параметров
- *nn\_models\_code.pdf* – код для работы с моделями нейросетей: поиск лучшей модели по результатам обучения, построение визуализаций
- *lin\_models\_code.pdf* – код для работы с линейной регрессией: обучение моделей, построение визуализаций, анализ коэффициентов
- *shap\_code.pdf* – код для выявления наиболее важных факторов в модели нейросети при помощи SHAP значений