

Содержание

1	Введение	2
2	Основная часть	3
3	Заключение	4

1 Введение

В настоящий момент ведутся работы по разработке и совершенствованию текущих возможностей сервиса. В период с 01.02.22 по 15.03.22 выполнены следующие задачи.

- Разработан модуль для загрузки, анализа и обработки собранных данных.

- Разработан модуль построения модели предсказания стоимости объектов недвижимости.

- Разработан модуль статистического анализа данных полученных в разные промежутки времени.

2 Основная часть

Для достижения поставленных целей НИОКР и всего проекта в целом в настоящий момент проводятся следующие работы.

Прежде всего разработан модуль загрузки, обработки и анализа данных. Модуль позволяет вычислить требуемые метрики, построить необходимые графики. Обработка данных состоит в поиске и обработке пропущенных значений, удалении выбросов, к категориальным признакам применяется one-hot-кодирование. Предобработанные данные в дальнейшем используются для построения всех требуемых моделей.

Начаты работы по разработке моделей предсказания стоимости объектов недвижимости. Выполнена преобработка данных по городу Москва, на основе данных построен набор моделей искусственного интеллекта, в дальнейшем на базе полученного набора построена единая модель - ансамбль моделей, при невысоком качестве каждой из моделей в отдельности единая модель позволяет обеспечить более высокие характеристики. К настоящему моменту достигнуты следующие показатели 83% объектов недвижимости из валидационной выборки предсказываются с погрешностью менее 10%, коэффициент детерминации R^2 равен 0.94. Добавлена возможность вычисления SHAP values, которые позволяют интерпретировать результаты вычисления моделей, оценивать важность каждого из признаков, описывающих объект недвижимости. Подход к построению моделей является обобщенным и впоследствии будет использоваться для построения моделей для других городов.

Начаты работы по разработке модуля, который будет анализировать и сравнивать данные, полученные в разные промежутки времени. Для города Москва вычислены необходимые метрики, построены требуемые графики. Выполнено сравнение посредством статистических тестов (тест Манна-Уитни, тест Колмогорова-Смирнова), тесты на уровне значимости 0.05 позволяют уверенно обнаружить изменения в данных.

3 Заключение

К настоящему моменту в результате проделанной работы можно сделать следующие выводы. Построенная модель предсказания стоимости объектов недвижимости не достигает заявленных показателей, 85% объявлений предсказывается с ошибкой менее 10%, в настоящий момент мы имеем 83%.

Последующие работы в период 15.03.22 - 30.04.22 будут направлены на увеличение количества используемых признаков, планируется добавить расстояние до метро, а также название станции метро и другие, что при использовании текущего модуля построения модели предсказания стоимости объектов недвижимости должно обеспечить заявленные показатели. Также будут проводиться работы по совершенствованию модуля статистического сравнения данных, полученных в разные промежутки времени, планируется добавить возможность сегментирования рынка и вычисления всех необходимых характеристик по соответствующим сегментам.

В дальнейшем усилия будут направлены на увеличение количества обрабатываемых городов. Планируется построить модели для города Санкт-Петербург и Нижний Новгород.