УДК 330.101.542

АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ГОРОДСКОГО РЫНКА ЖИЛЬЯ С ИСПОЛЬЗОВАНИЕМ СИГНАЛОВ ИНТЕРНЕТА[1]

А.В. Болдырева[1,2], О.А. Соболевский[1]

[1]*Московский физико-технический институт (государственный университет),*

[2]*Российская академия народного хозяйства и государственной службы при Президенте РФ*

anna.boldyreva@phystech.edu, oleg.sobolevskiy@phystech.edu

У статті показана можливість прогнозу цін на ринку нерухомості Москви на основі активності користувачів Інтернету. Моделі будуються з використанням методу групового обліку аргументів. Експерименти показали помилку 0.04%-5.88% для розрахунку поточної вартості нерухомості різного типу.
Ключові слова: пошукові запити, Інтернет, нерухомість,прогнозування,МГУА

The article shows the possibility of forecasting the prices on real estate market on the basis of the Internet users' activity. Models are built using the group method of data handling. The experiments showed the errors of 0.04%-5.88% for the current prices related to various types of property.
Keywords: search queries, internet, property, forecasting, GMDH

В статье показана возможность прогноза цен на рынке недвижимости Москвы на основе активности пользователей Интернета. Модели строятся с использованием метода группового учета аргументов. Эксперименты показали ошибку 0.04%-5.88% для расчета текущей стоимости недвижимости разного типа.
Ключевые слова: поисковые запросы, Интернет, недвижимость, прогнозирование, МГУА

1. Введение

1.1 Рынок недвижимости России, тенденции

После ипотечного кризиса 2008-го года на рынке недвижимости России сложилась не простая ситуация. Цены на квадратный метр в отношении к доходам населения показывают длительный спад, в отличие от ситуации в других странах. В 2012 году, на рынках недвижимости большинства стран произошел после-кризисный перелом тренда. В России, после незначительного подъема, снижение цен продолжилось. С другой стороны, рынки Австралии, Великобритании и Канады, очевидно, перегреты, что может стать предпосылкой нового кризиса, связанного с пузырями недвижимости[2] (рис. 1).

[1] Статья написана в рамках гранта программы «УМНИК», конкурс УМНИК-МФТИ-2016

[2] По данным международного портала «Экономист» http://www.economist.com/blogs/dailychart/2011/11/global-house-prices

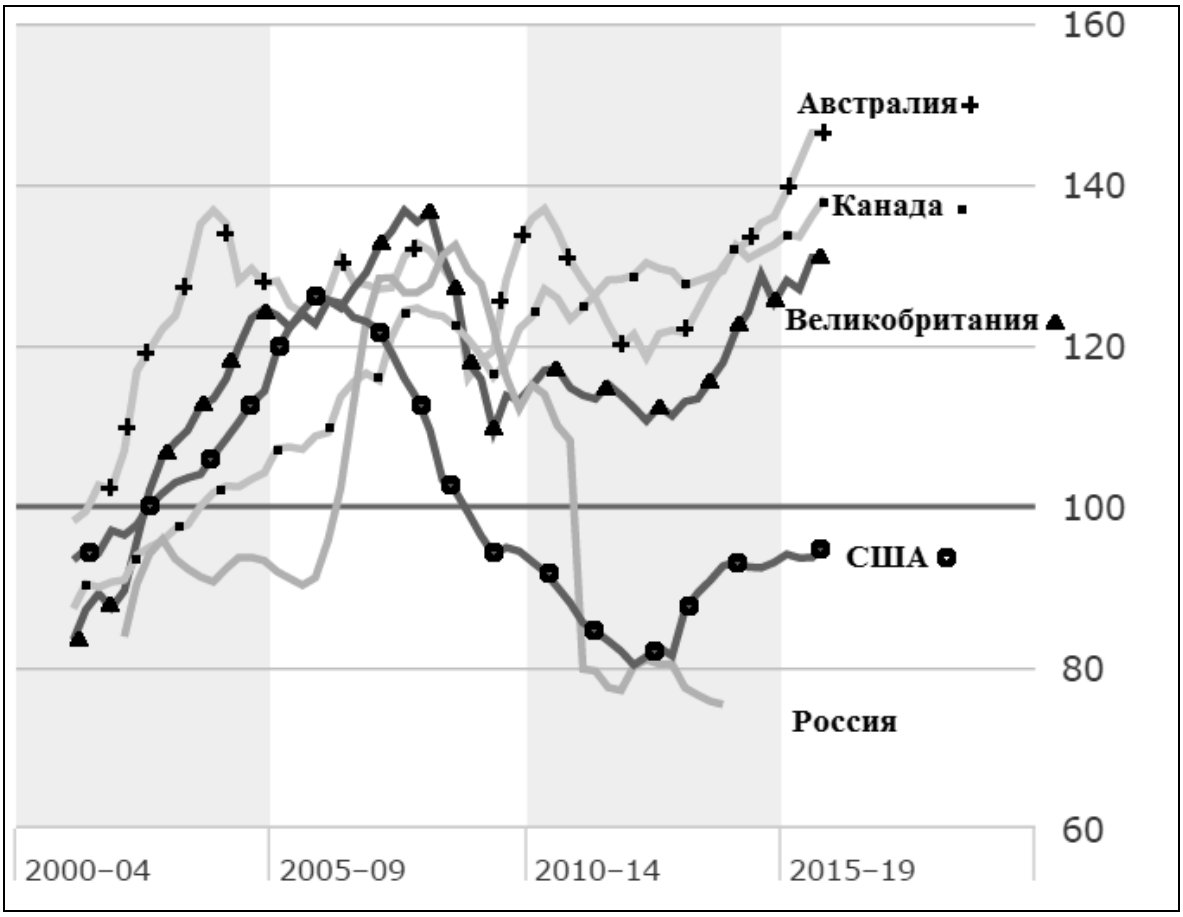

Рис. 1. Цены на недвижимость в России, США, Великобритании, Канаде, Австралии, относительно среднедушевого дохода, 2000-2015 гг.

На ценообразование жилой недвижимости влияет много факторов: общая экономическая ситуация в стране и платежеспособность населения, насыщенность рынка, снижение/повышение ипотечной ставки, изменение в жилищном законодательстве, смягчение/ужесточение ответственности продавцов и инвесторов, и др. Важное влияние оказывают цены на нефть. До 2012-го года, после кризиса, цены на нефть росли и затем снова начали падать. В 2014-м году следующий тренд падения — результат санкций. Но в России по-прежнему строится слишком мало жилья (менее 1 м^2 в год на человека[3]), чтобы ожидать дальнейшего длительного снижения цен. Также играет роль изменение требований к будущему жилью. Федеральные стандарты норм жилой площади на человека в России предполагают сейчас для семьи из трех человек не менее 18 м^2 на каждого[4]. Для сравнения, в Швеции на человека приходится в среднем 40 метров, в Германии 50 м^2, в США 70 м^2. Минимальный стандарт ООН 30 м^2 на человека. То есть, рынок жилья России еще очень далек от насыщения.

[3] http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/enterprise/building/, Федеральная служба Государственной статистики, официальная статистика, предпринимательство, строительство
[4] Жилищный Кодекс РФ, статья 38

В 2014-м году вышел закон об обязательном страховании ответственности застройщиков[5]. Закон оказывает дополнительное давление на цены, поскольку уменьшает риски вкладчиков[6].

Отдельно стоит рассмотреть вопрос ипотеки. В последнее время СМИ публикует очень много рекомендаций вкладываться в ипотеку. Какое влияние может оказать снижение ипотечных ставок на рынок недвижимости России? На рисунке 2 показан график индекса стоимости 1 квадратного метра недвижимости по Москве с 2001-го по 2016-й год[7]. Индекс представляет собой общерыночный показатель текущего среднего уровня цен на жилье, выраженный в рублях на квадратный метр.

Рис. 2. Индекс стоимости жилья руб/м^2, Москва

Посмотрим, как повлияли государственные программы обеспечения жильем основных групп населения. С 2004-го года начала действовать программа «Государственные жилищные сертификаты на 2004-2010 годы». С 2006-го года рост цен соответствует началу активной фазы целевой программы льготного ипотечного кредитования «Жилье». В 2007-м дополнительно стартовала программа «Молодая семья» и программа материнского капитала, и цены, после небольшого падения, продолжили рост, вплоть до начала кризиса 2008-го. Начиная с 2015-го года, ипотечные ставки начали постепенное

[5] https://rg.ru/2014/01/01/dolshiki-site.html, статья, Застройщиков обязали страховать ответственность

[6] http://realty.rbc.ru/news/57e287349a7947230f7a6a41, статья, Какую роль стали играть страховые компании в долевом строительстве

[7] http://www.irn.ru/gd/?class=all&type=1&period=0&Spear=week&grnum=1&currency=1&select=period, Цены на недвижимость и квартиры в Москве

снижение[8] — с 14.09% в июне 2015-го, до 12,54% в октябре 2016-го. На графике (рис. 2) мы наблюдаем соответствующий рост цен. Следовательно, появление дешевых денег на рынке приводит к удорожанию стоимости на жилье. Возможно, это происходит вследствие отставания ресурсной и инфраструктурной базы от потребностей рынка нового жилья.

1.2 Состояние вопроса и постановка задачи

Методы, традиционно используемые для прогноза цен на недвижимость, основаны преимущественно на тренде цен и индикаторов экономической конъюнктуры. В работе Георгия Стерника [1] предложена методика, которая базируется на результатах исследования эластичности цен, доходов населения и классификации рынков. Исследование подтвердило, что, при наличии прогнозов изменения душевых доходов населения, возможно и достаточно обоснованное прогнозирование цен на жилье. Однако официальная статистика государства публикуется очень неровно, и не всегда есть возможность применить авторегрессию или выявить тренд. В этом случае целесообразно применять нестандартные способы прогноза. Ник Макларен и Рачана Шанхог из Bank of England в своем исследовании [2] изучили динамику поисковой активности пользователей Интернета для анализа рынков труда и жилья Великобритании. Исследование показало, что с дальнейшим развитием этого метода прогнозирования, поисковые запросы Интернета станут важным инструментом экономического анализа. В статье, посвященной исследованию рынка недвижимости Лондона, ученые Эласдер Райя и Эбри Сенер [3] рассматривают территориальные особенности поиска жилья в столице Великобритании. Авторы анализируют данные популярного портала о недвижимости Великобритании, и выявляют закономерности поисковой активности жителей разных районов Лондона, внеся свой вклад в изучение пространственных особенностей города.

Анализ объемов продаж и цен на недвижимость с использованием поисковых запросов провели и Линн Ву и Эрик Бриньолфссон в своей работе [4]. В исследовании Николь Браун [5] изучалась возможность использования прогнозирования на рынке аренды жилья. В качестве индикатора был использован индекс S&P US REIT, в качестве вспомогательных сигналов для улучшения прогнозных моделей были собраны данные об объеме запросов в поисковую машину Google. Исследование подтвердило, что динамика запросов соответствует динамике рынка недвижимости с опережением в один шаг. Использование этой закономерности позволяет улучшить прогнозные модели.

В данной статье впервые применен метод исследования рынка недвижимости России на основе Интернет-активности пользователей Интернета. Мы рассматриваем разные категории покупателей недвижимости и проводим анализ округов Москвы с учетом разных задач потенциальных

[8] http://www.cbr.ru/statistics/?PrtId=ipoteka

инвесторов. В качестве таких задач выступают: покупка жилья для собственного проживания, инвестиции в недвижимость с целью получения дивидендов на рынке аренды, инвестиции с целью сохранения капитала.

Для построения объясняющих переменных в данном исследовании используется подход Михаила Столбова по формированию барометров финансовых показателей из поисковых запросов [6]. В настоящей работе прогноз проводится по модифицированному методу, предложенному ранее одним из авторов для построения прогностических моделей индикаторов экономической, социальной и финансовой конъюнктуры России [7-9].

2. Анализ данных

2.1 Термины

Диапазон данных. Мы рассматривали данные с октября 2014 года по ноябрь 2016 года, что вызвано ограничением предоставления данных по запросам сервисом Яндекса.

Индикаторы. В качестве индикаторов в работе использовалась статистика средних цен в рублях на кв. метр, полученная с помощью анализа реальных объявлений, размещенных в базах недвижимости ЦИАН[9], Авито[10], Домофонд[11]: средние цены на покупку и аренду квартир, комнат и нежилых помещений. Эти данные могут отличаться от официальных, так как не всегда официально проставляемые цены соответствуют реальной сумме, полученной за квартиру. Данные собирались специально разработанными для этой цели программами. В прогнозных моделях эти индикаторы брались в качестве зависимых переменных.

Дескрипторы. Дескриптор — часть слова, слово или словосочетание, служащие для формулировки запроса при поиске информации в поисковой системе. Для исследования были собраны базы дескрипторов — 150 тыс. слов и словосочетаний. Использовались как данные поисковых запросов статистического сервиса Яндекс[12], так и «упоминания» в сети. Данные, предоставляемые поисковыми машинами в части «упоминаний», отличаются от данных «поисковых запросов». Дескрипторы «запросов» включают в себя слова и словосочетания, которые пользователи набирают в поисковой строке. Дескрипторы «упоминаний» включают в себя данные о количестве Интернет-страниц Интернет-СМИ, материалы диалогов пользователей на различных форумах и социальных сетях, и др., где встречаются исследуемые слова и словосочетания, и публикуются на центральной странице Яндекса. В отличие

[9] http://www.cian.ru/
[10] https://www.avito.ru/rossiya/nedvizhimost
[11] http://www.domofond.ru/
[12] https://wordstat.yandex.com/

от данных поисковых запросов, данные «упоминаний» с течением времени меняются, поскольку поисковые роботы индексируют новые страницы, а старые страницы удаляются из выдачи. Поэтому ценность для исследователя представляют данные, скачанные по всему временному диапазону в короткий промежуток времени. Эти данные также собирались специально разработанными программами.

Топовая выборка дескрипторов и барометры — это элементы модели, которые использовались для решения прогнозных задач.

Топовая выборка дескрипторов — выборка динамик поисковых запросов, наиболее сильно коррелирующих с заданными индикаторами. Топовая выборка для каждого индикатора отбиралась из одного пула данных, это 150 тыс. дескрипторов.

Барометры — временные ряды, где значение в каждый период представляет собой среднее нормированное значение дескрипторов «топовой выборки» за соответствующий период. В прогнозной части работы барометры использовались в качестве объясняющих переменных.

2.2 Формирование топовой выборки и барометров

При формировании топовой выборки дескрипторы отбирались отдельно в рамках двух критериев:

- Отбор на основе коэффициента корреляции Пирсона (параметрический критерий, отражает степень линейной связи);
- Отбор на основе коэффициента ранговой корреляции Спирмена (непараметрический критерий, отражает близость трендов).

При оценке связи мы учитывали как положительную корреляцию, так и отрицательную корреляцию (анти-корреляция) с пороговыми значениями, соответственно (±)0.7. Дело в том, что при пороговом значении в (±)0.5 число отобранных дескрипторов резко возрастало до сотен переменных. При пороговом значении в (±)0.9 число дескрипторов резко падало до первых десятков. Поэтому выбранные значения (±)0.7 рассматривались, как некоторый компромисс.

При отборе дескрипторов рассматривалась не только связь одномоментных значений индикатора и дескрипторов, но и их связь при запаздывании реакции индикаторов на 1, 2 и 3 месяца.

В итоге для каждого индикатора было сформировано 16 топовых выборок с различиями в типе коэффициента корреляции (2 типа), знаке коэффициента корреляции (2 знака) и лаге (4 лага). Эти топовые выборки содержали 10-70 дескрипторов. На указанных топовых выборках были сформированы 16 барометров. Например, вот обозначения барометров для прогнозирования продаж квартир:

Apart.Sale_0+Pear; Apart.Sale_0–Pear; Apart.Sale_1+Pear; Apart.Sale_1–Pear; Apart.Sale_2+Pear; Apart.Sale_2–Pear; Apart.Sale_3+Pear; Apart.Sale_3–Pear; Apart.Sale_0–Spear; Apart.Sale_0+Spear; Apart.Sale_1–Spear; Apart.Sale_1+Spear; Apart.Sale_2–Spear; Apart.Sale_2+Spear; Apart.Sale_3–Spear; Apart.Sale_3+Spear.

В этих обозначениях цифры 0-3 указывают на лаг, знаки «+» и «–» отражают положительную и отрицательную корреляцию, слово Pear говорит о том, что барометр строится на основе коэффициента корреляции по Пирсону, слово Spear указывает на использование коэффициента ранговой корреляции по Спирмену.

2.3 Анализ распределения цен по округам

На рисунках 3-4 показана дифференциация округов Москвы в пределах МКАД, по ценам за квадратный метр по покупке/продаже и аренде в среднем за период с декабря 2013-го по октябрь 2016 по данным объявлений, скачанных с порталов ЦИАН, Авито, Домофонд. На представленной инфографике интенсивность окраски отражает цены за руб/м^2 для категорий:

- «Продажи. Комнаты», шкала [175000—240000], с шагом 4000 руб;
- «Продажи. Квартиры», шкала [165000—441000], с шагом 17000 руб;
- «Продажи. Нежилые», шкала [171000—417000] , с шагом 15000 руб;
- «Аренда. Комнаты», шкала [850—1035] , с шагом 10 руб;
- «Аренда. Квартиры», шкала [810—1400], с шагом 40 руб;
- «Аренда. Нежилые», шкала [18500—63000] , с шагом 2500 руб.

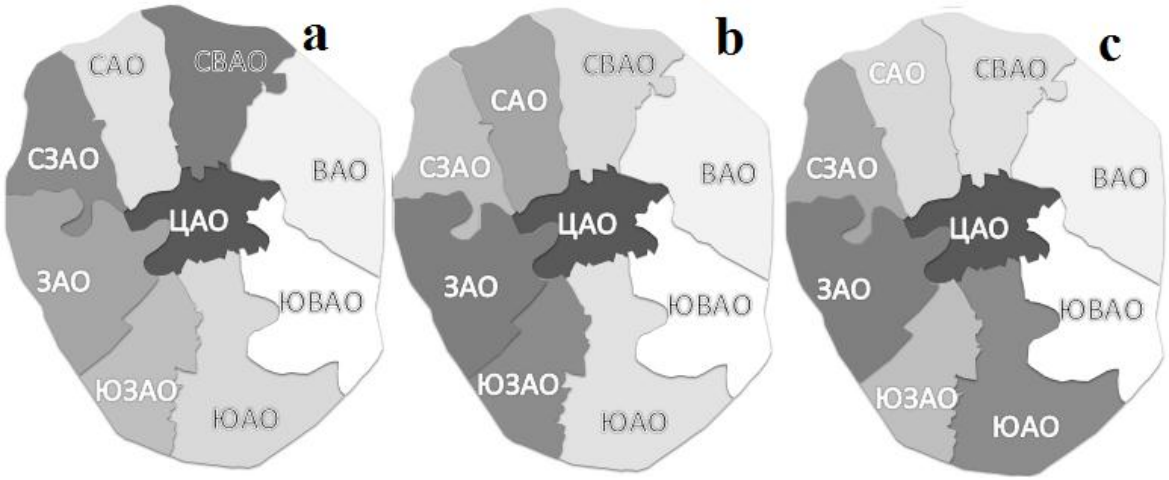

Рис. 3 Продажа, a — комнаты, b — квартиры, c — нежилые помещения

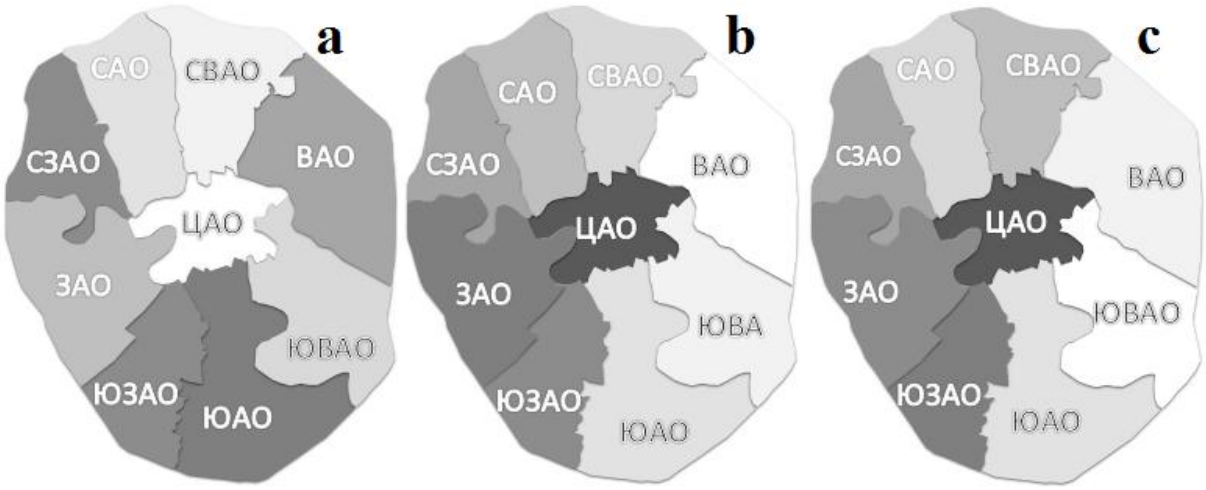

Рис. 4 Аренда, a — комнаты, b — квартиры, c — нежилые помещения

Восточные районы для аренды и продажи предлагают преимущественно наименьшую цену за квадратный метр, центр (ЦАО — Центральный Административный Округ) – самую высокую. Для центра исключением являются очень низкие цены на аренду комнат. В СВАО (Северо-Восточный) и ЗАО (Западный АО) можно также арендовать комнаты с меньшими затратами, чем покупать, в сравнении с другими округами. Южный АО выделяется высокими ценами на аренду комнат и на покупку нежилых помещений.

Интересно проанализировать, как изменялись цены в динамике. В качестве примера рассмотрим цены на покупку и аренду квартиры.

При сравнении цен на покупку (табл. 1) и аренду (табл. 2), очевидно, что вложение средств в расчете на повышение стоимости недвижимости целесообразнее в Центральном округе, а арендовать выгоднее всего в Восточном округе.

Таблица 1

Квартиры, продажа, кв. метр, объявления, округа Москвы

Кв, прод.	дек.13	1 кв.14	2 кв.14	3 кв.14	4 кв.14	1 кв.15	2 кв.15	3 кв.15	4 кв.15	1 кв.16	2 кв.16	3 кв.16	окт.16
ВАО	164862	166806	170776	173000	180708	184852	183337	176474	172368	172180	170014	168314	167346
ЗАО	205935	212091	224907	230505	239964	243489	239690	233644	240533	255642	244865	257186	248097
САО	185730	188757	196219	202372	211404	217322	214195	204100	201379	203636	203601	205235	203626
СВАО	175670	178371	184139	187706	197709	200610	199475	193775	186637	184598	180092	179790	182404
СЗАО	185160	188548	198723	204522	213342	215504	211415	202566	197638	201240	198594	208490	201195
ЦАО	311325	321166	340948	353343	373063	421987	416828	451668	531452	588381	534036	533189	546908
ЮАО	166795	169006	174041	177730	184107	185421	185169	179946	174327	172886	173450	174261	172650
ЮВАО	156608	158412	162391	164904	173219	175965	173549	167713	162211	160218	160200	159229	158151
ЮЗАО	195738	198634	205560	210845	218565	224750	226391	216787	211836	216255	210969	217518	215980

Таблица 2

Квартиры, аренда, кв. метр, объявления, округа Москвы

Кв.Аренда	дек.13	1 кв.14	2 кв.14	3 кв.14	4кв.14	1кв.15	2кв.15	3кв.15	4кв.15	1кв.16	2кв.16	3кв.16	4кв.16
ВАО	862.08	846.81	825.69	831.66	844.1	841.28	809.57	792.29	800.02	780.21	771.16	777.74	787.73
ЗАО	942.82	945.18	926.98	944.51	978.8	977.98	946.12	942.67	955.33	944.95	906.38	928.41	940.13
САО	888.32	890.36	872.3	881.12	902.45	900.89	859.39	847.35	880.65	854.22	835.37	835.14	862.34
СВАО	883.81	881.31	856.16	870.28	896.22	881.54	846.27	826.47	834.43	825.5	812.36	813.62	830.98
СЗАО	897.75	894.21	878.17	884.95	914.79	912.91	869.68	863.49	866.76	849.69	852.9	853.27	880.28
ЦАО	1258	1277	1305.4	1324	1344.8	1420.7	1463.9	1524.2	1530	1489	1413.7	1370.1	1377.1
ЮАО	883.57	882.92	864.66	869.21	892.28	881.03	830.57	822.75	836.39	810.44	798.7	809.28	828.54
ЮВАО	854.06	846.3	822.17	837.45	870.87	856.89	809.47	788.53	786.38	777.17	774.04	792.56	803.18
ЮЗАО	916.37	911.59	894.79	903.09	932.7	917.52	880.58	873.92	885.09	858.23	847.81	863.45	877.47

Рассчитаем срок окупаемости (табл. 3) в первом приближении, без учета возможного роста цен на недвижимость, коммунальных платежей, и ремонта квартиры, по формуле:

$$P=S/(R_m*12)$$

где P — окупаемость в годах (payback), S — цена за квадратный метр продажи/покупки (sell), R_m— цена аренды квадратного метра в месяц (rent).

Таблица 3

Окупаемость вложений в покупку квартиры по округам

Окуп-сть	дек.13	1 кв.14	2 кв.14	3 кв.14	4кв.14	1кв.15	2кв.15	3кв.15	4кв.15	1кв.16	2кв.16	3кв.16	4кв.16
ВАО	15.9	16.4	17.2	17.3	17.8	18.3	18.9	18.6	18.0	18.4	18.4	18.0	17.7
ЗАО	18.2	18.7	20.2	20.3	20.4	20.7	21.1	20.7	21.0	22.5	22.5	23.1	22.0
САО	17.4	17.7	18.7	19.1	19.5	20.1	20.8	20.1	19.1	19.9	20.3	20.5	19.7
СВАО	16.6	16.9	17.9	18.0	18.4	19.0	19.6	19.5	18.6	18.6	18.5	18.4	18.3
СЗАО	17.2	17.6	18.9	19.3	19.4	19.7	20.3	19.5	19.0	19.7	19.4	20.4	19.0
ЦАО	20.6	21.0	21.8	22.2	23.1	24.8	23.7	24.7	28.9	32.9	31.5	32.4	33.1
ЮАО	15.7	16.0	16.8	17.0	17.2	17.5	18.6	18.2	17.4	17.8	18.1	17.9	17.4
ЮВАО	15.3	15.6	16.5	16.4	16.6	17.1	17.9	17.7	17.2	17.2	17.2	16.7	16.4
ЮЗАО	17.8	18.2	19.1	19.5	19.5	20.4	21.4	20.7	19.9	21.0	20.7	21.0	20.5

Период окупаемости растет во всех округах, соответственно, доходность вложений в недвижимость Москвы падает. Коммерческое вложение в недвижимость (квартиры) с целью получения дивидендов, целесообразнее на данный момент в ЮВАО, ЮАО и ВАО. Но, чтобы анализ был полным, нам необходим расчет с учетом прогноза вперед, как будут меняться цены на покупку/продажу квартиры, на аренду, и, соответственно, на окупаемость. Для этого мы изучим данные поисковой активности пользователей Интернета.

2.4 Анализ активности пользователей Интернета

В таблицах 4-5 показана динамика поисковых запросов в период 4 квартал 2014 – октябрь 2016-го. Представлены средние относительные значения числа запросов в месяц. Для каждого округа был составлен пул поисковых запросов в соответствии с категориями. Например, для анализа покупательской активности пользователей Интернета относительно продажи/покупки квартиры, пул запросов содержал дескрипторы: квартира купить, квартира без посредников, цена недвижимость, жилье квартира купить, авито квартиры, продажа квартир, агентство недвижимости квартиры, стоимость квартир, сколько стоит моя квартира, продать квартиру, купить квартиру, циан квартира продавать, продажа недвижимости, недвижимость и цены, квартира покупка, стоимость недвижимости, недорогие квартиры, вторичное в жилье купить, квартиру дешево, квартира студия, купить студию, коммуналка купить, вторичка купить, вторичное жилье, недвижимость официальный сайт, новостройка, застройщик квартира, застройщик новостройка, квартиры с отделкой, новые квартиры, новое жилье купить, цены от застройщика и т.д.

Всего было использовано около 100-200 дескрипторов для каждой категории. Дескрипторы отбирались с помощью сервиса Яндекса для поисковых запросов[13] в разделе «Что искали со словом «категория». Кроме того использовался сервис на основной странице Яндекса «вместе с этим словом ищут:…».

Таблица 4

Продажи квартир и комнат, Москва, дескрипторы, 2014-2016

Кв.Комн.Прод	4кв.14	1кв.15	2кв.15	3кв.15	4кв.15	1кв.16	2кв.16	3кв.16	4кв.16
ВАО	117.36	103.79	96.737	127.92	130.59	159.74	146.92	155.96	168.18
ЗАО	94.617	81.832	72.116	113.99	114.03	141.24	136.54	146.91	154.78
САО	64.953	60.452	54.701	76.776	76.072	96.538	95.365	105.32	118.35
СВАО	82.268	76.997	66.534	96.2	97.1	118.82	113.33	126.02	138.64
СЗАО	46.151	46.046	39.192	53.906	54.98	69.553	66.095	71.474	77.518
ЦАО	90.589	91.757	81.003	94.891	103.41	112.29	101.26	115.66	122.65
ЮАО	140.44	127.08	107.23	158.02	161.82	198.53	178.92	203.91	228.12
ЮВАО	103.92	96.75	86.263	118.22	123.44	147.44	132.01	147.81	158.77
ЮЗАО	162.92	165.21	131.74	98.802	160.15	183.2	138.81	153.6	222.65

[13] https://wordstat.yandex.com/

Таблица 5

Аренда квартир и комнат, Москва, дескрипторы, 2014-2016

Кв.Комн.Арен	4кв.14	1кв.15	2кв.15	3кв.15	4кв.15	1кв.16	2кв.16	3кв.16	4кв.16
ВАО	88.774	86.495	77.565	126.96	117.85	125.01	134.32	171.62	167.61
ЗАО	77.53	73.498	63.757	108.09	99.046	105.95	109.46	149.87	144.55
САО	60.194	62.226	59.822	91.542	77.717	88.409	94.184	121.27	116.77
СВАО	67.49	63.795	56.901	99.619	92.219	98.824	105.1	131.92	138.59
СЗАО	33.347	36.218	33.607	49.793	43.353	45.582	57.86	66.543	67.391
ЦАО	47.29	44.44	38.293	64.353	60.039	65.976	66.598	93.589	99.688
ЮАО	99.032	96.548	81.766	134.99	128.8	130.41	151.74	226.9	227
ЮВАО	81.016	80.966	73.59	112.95	106.83	115.14	119.39	165.51	168.53
ЮЗАО	162.92	165.21	131.74	98.802	160.15	183.2	138.81	153.6	222.65

По динамике можно предположить, что растет оживление на рынке недвижимости Москвы. Мы предполагаем, что это связано со снижением ипотечной ставки в июле 2015-го года. На этот факт мы указывали в разделе «Введение». Интересно, что в Центральном административном округе имеет место малая активность, связанная с покупкой/продажей/арендой квартир и комнат по сравнению с другими округами, что коррелирует с самым высоким сроком окупаемости недвижимости в этом районе Москвы. Наиболее популярны у пользователей Интернета южные округа, наименее — Северо-Западный АО, хотя и там наблюдается рост поисковой активности.

2.5 Достоверность данных

Можно поставить вопрос: насколько достоверна информация, полученная на основе изучения сетевой активности пользователей Интернета?

Мы не смогли подтвердить исследование Николь Браун [5] для рынка Москвы, поскольку в России пока не сформирована в должной степени культура открытых данных с регулярной публикацией реальной статистики оформления собственности.

Но мы можем рассмотреть, например, возможность анализа поисковых запросов, связанных с ипотекой в сравнении с индикатором ипотечного кредитования, предоставляемым Центральным Банком РФ[14] (рис. 5). Для группы поисковых запросов в категории «Ипотека», были выбраны дескрипторы, содержащие слова и словосочетания: ипотека, взять ипотеку, ипотека [название района], ипотека в [банк], калькулятор ипотеки, квартира в ипотеку, условия ипотеки и др…

[14] http://www.cbr.ru/statistics/?PrtId=ipoteka

Рис. 5 Динамика ипотечных кредитов и поисковых запросов по ипотеке, данные нормированы

Мы видим, что динамика дескрипторов в категории «Ипотека» в целом соотносится с динамикой заключения договоров на ипотечное кредитование. Значение корреляции достаточно высокое — 0.71. С учетом резкого скачка с началом 3-го квартала, можно предположить, что к концу 2016 года количество договоров превысит уровень 2014-го.

Другим примером может служить анализ поисковых запросов, связанных с обеспеченностью жителей районов Москвы. Индикатором являются данные Сбербанка России[15].

Для группы поисковых запросов в категории «Обеспеченность» были выбраны дескрипторы, содержащие слова и словосочетания: элитные, бизнес-класс, купить доллары, курс доллара, купить дорогие, ювелирные магазины, такси, купить авто, аренда авто, отель, заказать ресторан и др. Сравним квартальную динамику зарплат по Москве со средней динамикой поисковых запросов, группа «Обеспеченность» (рис. 6).

[15] http://www.sberbank.com/ru/opendata Открытые данные

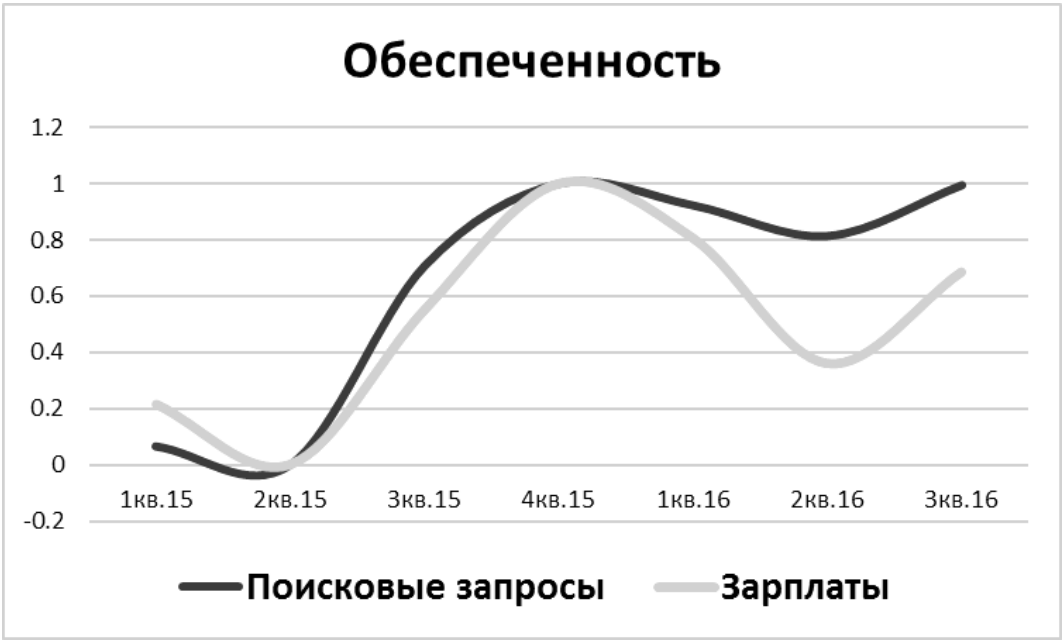

Рис. 6 Сравнение динамики средней зарплаты по Москве
с интенсивностью поисковых запросов, группа «Обеспеченность»,
данные нормированы

Мы видим очевидное соответствие обеих динамик. Это подтверждает и уровень корреляции между ними 0.88. При этом мы так же можем сделать вывод, что покупательная активность длится еще некоторое время после периода высоких зарплат.

Для проверки правильности выбора группы дескрипторов был проведен анализ сравнительной динамики по временной шкале всех исследуемых округов в регионе. На рисунке 7 представлены данные поисковых запросов группы «Обеспеченность» в динамике с 4-го кв. 2014 по 4-й кв. 2016-й года[16].

Округа ЦАО и САО показывают самый высокий уровень интереса пользователей Интернета. Округ ЮАО не соответствует общему тренду. Можно предположить, что причиной такого расхождения являются «серые зарплаты» в этом округе. Для более точного анализа динамики благосостояния и возможных источников заработков жителей округов нужны значения индикаторов конкретно по каждому округу.

[16] Стоит учесть, что это относительные данные, умноженные на 1млн, то есть это данные, относительно общего количества поисковых запросов, а не абсолютные цифры.

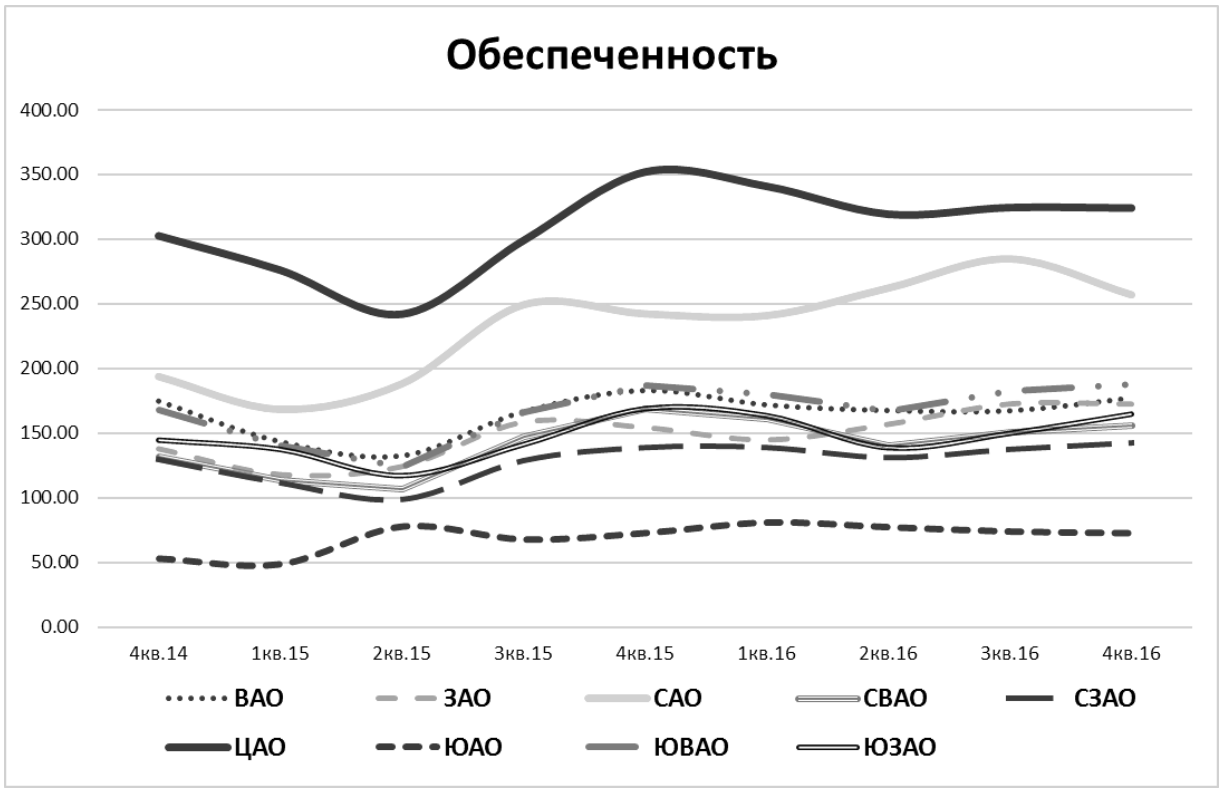

Рис. 7 Динамика средних показателей поисковых запросов группы «Обеспеченность» по округам Москвы

3. Прогнозирование

3.1 Среда моделирования

В исследованиях поисковых запросов очень большое внимание уделяется процедуре прогнозирования, ведь высокий уровень точности прогноза является дополнительным обоснованием применения сигналов Интернета.

Для анализа покупательской активности мы использовали очевидные дескрипторы, такие как: купить квартиру, квартиры от застройщика, недорогая квартира, и т.д. Но процессы, служащие триггером цен на рынке недвижимости, могут быть неочевидными, связанными с экономикой и политикой в масштабе страны. Поэтому для прогнозирования была взята большая база в 150 тыс. дескрипторов, на которую мы ссылались выше.

Построение прогнозных моделей есть результат моделирования. Мы строили прогнозные модели, которые связывали уровни цен (индикаторы) и интенсивность поисковых запросов (значения барометров). Для прогнозирования применялся метод группового учета аргументов (МГУА) [10], реализованный в оболочке GMDH Shell, или, сокращенно, GS [11]. Целью

экспериментов было не только построить наилучшие прогнозные модели, но и продемонстрировать возможности МГУА в рамках GS.

3.2 Алгоритмы

Все алгоритмы МГУА направлены на поиск моделей оптимальной сложности, которая понимается как наилучший компромисс между поведением модели на тестовой и проверочной выборках. Поиск модели ведется в заданном классе моделей, которым в оболочке GS являются полиномиальные модели.

В работе использовались два алгоритма построения моделей из GS: комбинаторный и нейросетевой. Оба алгоритма можно интерпретировать как «многорядные», однако понятие термина «ряд» в них разное. В комбинаторном алгоритме каждый ряд или этап отражает сложность модели, т.е. число включенных в нее независимых переменных, и число этапов всегда конечно. В нейросетевом алгоритме ряд отражает очередной шаг итерационного процесса селекции (теоретически бесконечного). Используя терминологию нейросетей, в литературе такие ряды называют слоями. Перебор моделей в комбинаторном алгоритме и формирование моделей в нейросетевом алгоритме прекращаются, когда выполнены требования выбранного критерия качества модели [10].

В комбинаторном алгоритме все модели распределены по этапам, и в процессе перебора производится проверка всех моделей данного этапа, а затем выполняется переход к следующему. В таком алгоритме исключена потеря лучшей модели, и модель представляется одним уравнением, связывающим в нашем случае индикатор и барометры. Недостаток алгоритма заключается в большом времени вычислений.

При проведении моделирования с комбинаторным алгоритмом, из-за ограничений по времени счета использовалось лишь простые линейные полиномы барометров с добавлением их квадратных корней. Последние были включены в модели для учета зависимости, ослабевающей с ростом числа запросов – например, для случая двух переменных-барометров имеем:

$$y = w_{00} + w_{11}x_1 + w_{12}\sqrt{x_1} + w_{21}x_2 + w_{22}\sqrt{x_2}\ .$$

Здесь y — это индикаторы (цены), x_i — барометры, w_{ij} — коэффициенты при объясняющих переменных.

Нейросетевой алгоритм — это полиномиальная нейронная сеть, где на текущем слое переменными являются выходы лучших регрессионных моделей предыдущего слоя. Из каждой пары таких переменных формируются модели данного слоя, из которых выбираются лучшие для перехода на следующий слой. Здесь возможна потеря лучшей модели в процессе селекции, однако время счета резко сокращается. Недостаток нейросетевого алгоритма заключается в сложности представления модели в форме иерархической системы уравнений и в неочевидности влияния отдельных барометров.

При проведении моделирования с нейросетевым алгоритмом были приняты следующие настройки: ограничение на начальную ширину слоя 5, максимальное количество слоев 6. Начальная ширина слоя — это количество лучших моделей, которые передаются со слоя на слой. Здесь модели ищутся в виде квадратичного полинома с добавлением квадратных корней барометров. Таким образом, в этих моделях была возможность учесть и усиливающуюся, и ослабевающую с ростом запросов зависимость цен от барометров. Были также учтены связи переменных за счет включения в модель их попарных произведений:

$$y = w_0 + w_1x_1 + w_2\sqrt{x_1} + w_3x_1^2 + w_4x_2 + w_5\sqrt{x_2} + w_6x_2^2 + w_7x_1x_2 + w_8x_1\sqrt{x_2} + w_9x_2\sqrt{x_1} + w_{10}\sqrt{(x_1x_2)}\ .$$

Здесь y — это индикаторы (цены), x_i — барометры, w_j — коэффициенты при объясняющих переменных.

3.3 Процесс моделирования

Мы проверяли возможности двух указанных выше алгоритмов, комбинаторного и нейросетевого, в 2-х вариантах:

- в первом варианте в стартовый набор переменных включались все 16 барометров, которые были описаны выше в п. 2.2. В этом наборе учитывались значения дескрипторов с лагами, равными 0, 1, 2 и 3 месяца.
- во втором варианте в стартовый набор переменных дополнительно включались значения индикатора с запаздыванием до 3-х месяцев.

Прогноз на несколько месяцев в системе GS выполняется на основе не одной, а нескольких моделей, рассчитанных на разные горизонты прогноза. Рассмотрим в качестве примера прогноз на 3 месяца. Тогда строятся сразу 3 прогнозные модели: на 1, 2 и 3 месяца. Очевидно, что модели, в общем случае, оказываются разными. Пусть мы находимся в момент времени t. Тогда расчетная схема выглядит следующим образом:

- для прогноза на момент t+3 используется расчетное значение по одной модели, которое и принимается в качестве прогноза;
- для прогноза на момент t+2 используются расчетные значение по двум моделям, а именно, по одной модели, ориентированной на горизонт 3 месяца, и по другой модели – для горизонта 2 месяца. В качестве прогноза берется среднее этих двух значений;
- для прогноза на момент t+1 используются расчетные значение по трем моделям, а именно, на одной, ориентированной на горизонт 3 месяца, по другой – для горизонта 2 месяца, и по третьей – для горизонта 1 месяц. В качестве прогноза берется среднее этих трех значений.

Описанная расчетная схема суть эвристика, которая не имеет теоретического обоснования. Однако, по словам главного конструктора GS, она стабилизирует прогноз намного лучше, чем другие подходы, которые были тестированы разработчиками GS.

Именно по такой схеме рассчитывался прогноз на 6 месяцев и 3 месяца, результаты которого представлены ниже.

Для «валидации» модели, т.е. оценки ее качества, использовалась перекрестная проверка, которая в английской литературе называется *k*-fold cross validation. Здесь все множество точек наблюдений разбивается на *k* частей. На одной части модель обучается, а на других (*k*–1) частях модель проверяется и фиксируется полученная ошибка. Эта процедура повторяется *k* раз с разными обучающими выборками. Рассчитывается средняя ошибка, полученная на всех *k* шагах эксперимента. В нашей работе мы использовали проверку при *k*=2, то есть на двух наборах. В данном случае примененный способ есть не что иное, как реализация известного симметричного критерия регулярности [10].

В качестве оценки построенной модели применялась средняя абсолютная процентная ошибка (Mean Absolute Percentage Error):

$$MAPE = 1/N \sum_t \left| (Y_t - y_t)/y_t \right| * 100\%$$

Здесь: y_t — измеренное значение индикатора в момент времени t, Y_t — прогнозное значение индикатора в момент времени t.

3.4 Результаты моделирования

Как уже указывалось выше, мы проверяли возможности комбинаторного и нейросетевого алгоритмов с использованием и без использования в стартовом наборе запаздывающих значений индикатора наряду со значениями барометров. Модель, где использовались запаздывающие значения индикатора, назовем в статье моделью с авторегрессией. При проверке использовались прогнозы на 3 и 6 месяцев, чтобы изучить возможность краткосрочного и среднесрочного прогнозирования на основе одного и того же набора барометров. Результаты моделирования последовательно представлены в таблицах 6-9. Как оказалось, качество моделей с авторегрессией практически совпадали с качеством моделей без авторегрессии, поэтому результаты прогноза по таким моделям не были включены в указанные таблицы.

Таблица 6

Значения MAPE при прогнозировании на 6 месяцев

	нейросеть	комбинаторный
Мск.кв.продажа	0.04%	0.28%
Мск.кв.аренда	0.07%	0.31%

Продолжение таблицы 6

Мск.комн.продажа	0.07%	0.48%
Мск.комн.аренда	0.02%	0.12%
Мск.нежил.продажа	0.34%	1.53%
Мск.нежил.аренда	0.19%	1.14%

Таблица 7

Значения MAPE при прогнозировании на 6 месяцев на выборке, которая не использовалась для построения модели

	нейросеть	комбинаторный
Мск.кварт.продажа	3.59%	3.34%
Мск.кварт.аренда	5.88%	2.93%
Мск.комн.продажа	5.40%	3.31%
Мск.комн.аренда	2.24%	2.02%
Мск.нежил.продажа	15.28%	14.23%
Мск.нежил.аренда	15.82%	12.65%

Несмотря на хорошие значения MAPE в таблице 6, отличие реальных значений от прогнозных для нежилых помещений (табл. 7) превышает 10% (выделено серым цветом). Это можно объяснить тем, что возможности прогноза с использованием данных Интернет-активности пользователей существенно лучше для более популярных типов недвижимости, таких, как квартиры и комнаты, чем для нежилых помещений, которыми интересуется сравнительно меньшая прослойка населения. Поэтому для краткосрочного прогнозирования на три месяца были выбраны только комнаты и квартиры. Результаты представлены в таблицах 8 и 9.

Таблица 8

Значения MAPE при прогнозировании на 3 месяца

	нейросеть	комбинаторный
Мск.кварт.продажа	0.19%	0.54%
Мск.кварт.аренда	1.12%	0.37%
Мск.комн.продажа	0.10%	0.37%
Мск.комн.аренда	0.05%	0.38%

Таблица 9

Значения MAPE при прогнозировании на 3 месяца на выборке, которая не использовалась для построения модели

	нейросеть	комбинаторный
Мск.кв.продажа	0.69%	1.13%
Мск.кв.аренда	1.71%	1.97%
Мск.комн.продажа	3.17%	3.77%
Мск.комн.аренда	0.43%	0.79%

Прогнозируемость для 3-х месяцев улучшилась и составила 0.4%-3.8% для квартир и комнат.

На рисунках 8 и 9 приведены графики динамики продаж квартир и аренды комнат Москвы.

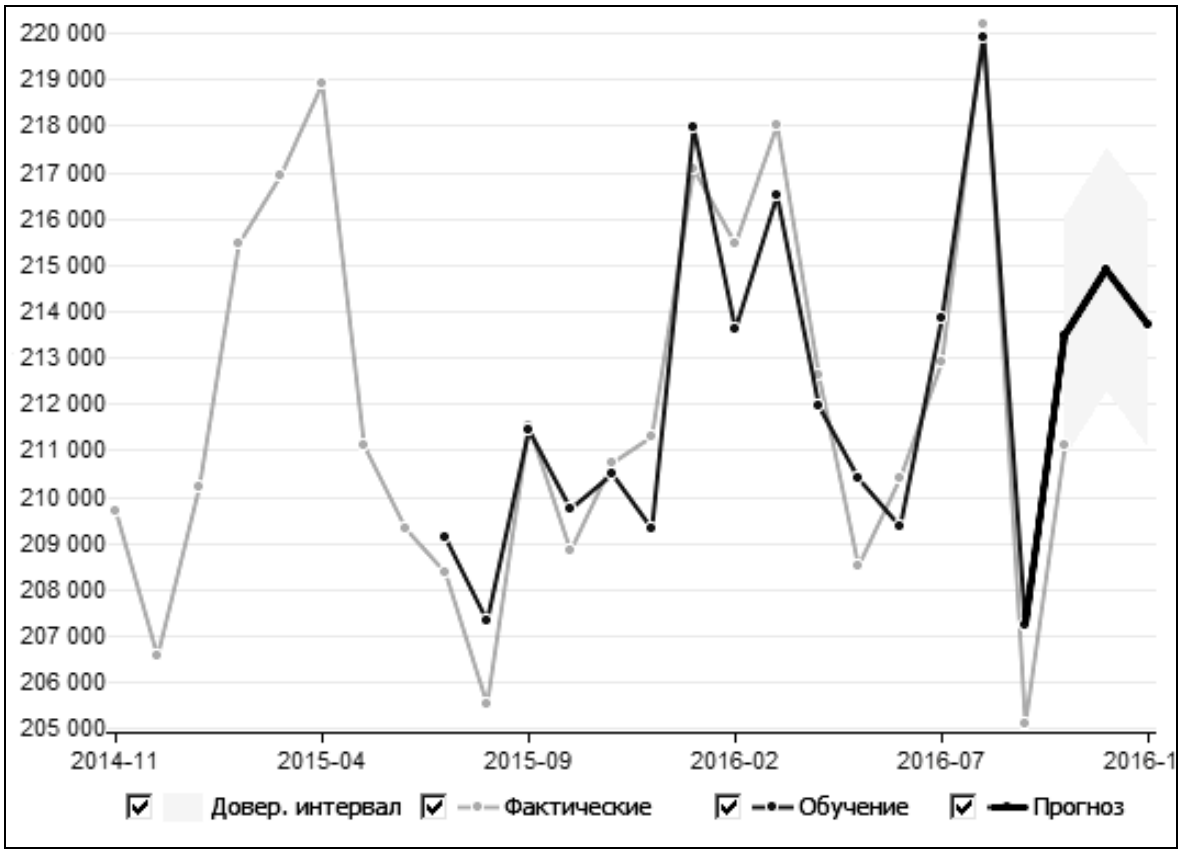

Рис. 8 Продажа квартир, комбинаторный алгоритм, прогноз на 3 месяца

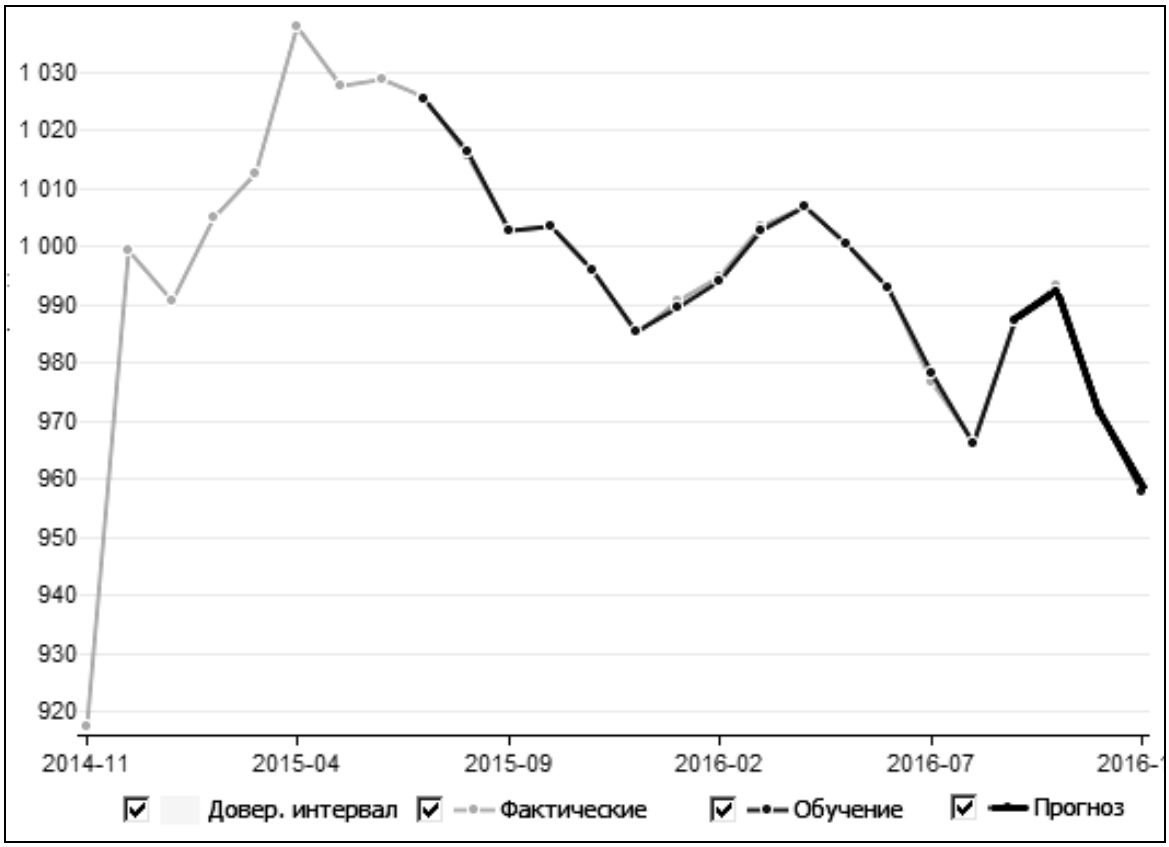

Рис. 9 Аренда комнат, нейросетевой алгоритм,
модель с авторегрессией, прогноз на 3 месяца

Приведем уравнения прогнозных моделей на 1, 2 и 3 месяца для цен на продажу квартир. Все модели получены в рамках комбинаторного алгоритма с использованием авторегрессии

Месяц: +1

$$Prices[t] = 232514.0 - 27441.6 * \sqrt{Apart.Sale\_2 - Spear[t-2]} - 6725.3 * Apart.Sale\_1 - Pear[t-3]$$

Месяц: +2

$$Prices[t] = 232214.0 - 26031.4 * \sqrt{Apart.Sale\_2 - Pear[t-2]} - 8595.2 * Apart.Sale\_3 - Pear[t-2]$$

Месяц: +3

$$Prices[t] = 190626.0 + 20819.3 * \sqrt{Apart.Sale\_1 - Pear[t-1]} + 25503.3 * Apart.Sale\_3 + Spear[t-3]$$

Описания барометров в приведенных формулах дано выше в п. 2.2.

Заключение

Результаты расчетов показывают, что цены на недвижимость в ближайшее время будут волатильны, с тенденцией к снижению. Для более значимого анализа нужно увеличить диапазон исследования и провести прогнозирование с шагом в один год, а не месяц, как это было сделано в работе.

Информация данного исследования может быть полезна:

- Покупателям жилья для собственного проживания;
- Инвесторам, вкладывающим свободные средства в недвижимость в целях сбережения капитала от инфляции;
- Инвесторам, вкладывающим средства в недвижимость с целью получения дохода на рынке аренды жилья.

Исследование показало, что поисковые запросы применимы для прогнозирования цен на некоммерческую недвижимость в Москве. Можно сделать вывод, что для прогноза цен на недвижимость в России целесообразно вместе с обычными методами применять динамики поисковых запросов. Модели, описанные в статье, дают ошибку прогнозирования в 0.04-5.88%.

Литература

1. Стерник Г.М. Методика прогнозирования цен на жилье в зависимости от типа рынка / Г.М. Стерник // Имущественные отношения в РФ. – 2011. – №1 (112). – С. 43-47.
2. McLaren N., Shanbhogue R., Using Internet Search Data as Economic Indicators // Bank of England Quarterly Bulletin. – 2011. – № 11, Q2. – P. 134-140.
3. Raea A., Sener S., How website users segment a city: The geography of housing search in London // Cities. – 2016. – V. 52. – P. 140-147.
4. Wu L., Brynjolfsson E., The future of prediction: How Google searches foreshadow housing prices and sales // Economic Analysis of the Digital Economy. – 2015. – P. 89-118.
5. Braun N., Google search volume sentiment and its impact on REIT market movements // Journal of Property Investment & Finance. – 2015. – V. 34, I.3. – P. 249- 262.
6. Столбов М.И. Статистика поиска в Google как индикатор финансовой конъюнктуры / М.И. Столбов // Вопросы экономики. – 2011. – № 11. – С. 79-84.

7. Болдырева А.В. Прогноз событий экономической направленности по запросам в Интернет / А.В. Болдырева, С.А. Маруев, А.В. Трусов (ред.) // Разработка моделей и методов анализа социально-технической среды бизнеса, годовой отчет лаборатории мат. методов анализа социальных сетей, РАНХиГС, Россия. – 2014. – С. 128-134.

8. Болдырева А.В. Применение метода МГУА на основе интенсивности поисковых запросов в сети Интернет для прогноза рынка недвижимости/ А.В. Болдырева // Труды Второй молодежной научной конференции «Задачи современной информатики». – 2015. – С. 46-51.

9. Болдырева А.В. Построение прогнозных моделей экономической и социальной конъюнктуры по интенсивности запросов в поисковой сети Интернет / А.В. Болдырева // Современная экономика: теория, политика, инновации. Сборник студенческих научных работ. – Москва: РАНХиГС, 2016. – С. 36-61.

10. Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования. – Киев: Наук. думка, 1985. – 216 с.

11. GMDH Shell, https://www.gmdhshell.com/