



ИНТЕЛЛЕКТУАЛЬНАЯ МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ПРОДАЖ НЕДВИЖИМОСТИ НА ПЕРВИЧНОМ РЕГИОНАЛЬНОМ РЫНКЕ

Коробкова Ольга Викторовна

старший преподаватель,
Южно-Уральский государственный университет,
Челябинск, Россия
E-mail: ufimtcevaov@susu.ru

Предмет исследования: массив данных, включающий социально-экономические параметры транзакций на рынке первичного жилья Челябинской области.

Цель исследования: создать нейросетевую модель для прогнозирования объемов продаж недвижимости с временным горизонтом в один год на основе ретроспективных данных. Прогностическая модель должна обладать простотой настройки и эксплуатации, а также предусматривать наличие интуитивно понятного пользовательского интерфейса.

Методы исследования: методология анализа данных CRISP-DM, корреляционный анализ, нормализация, k-fold кросс-валидация, построение полносвязной нейронной сети с двумя скрытыми слоями и сигмоидальной функцией активации, метод оптимизации L-BFGS.

Объект исследования: прогнозирование объемов продаж на первичном региональном рынке жилищного строительства.

Основные результаты исследования: создана нейросетевая модель прогнозирования с использованием аналитической платформы Loginom Community, обладающая модульной структурой, что обеспечивает ее универсальность. Модель успешно прошла тестирование на валидационной выборке, продемонстрировав точность прогнозирования в 88,33 %.

Ключевые слова: прогнозирование, перцептроны, анализ данных, принятие решений, временные ряды, алгоритмы обучения нейронных сетей.

AN INTELLIGENT MODEL FOR FORECASTING REAL ESTATE SALES IN THE PRIMARY REGIONAL MARKET

Olga V. Korobkova

Senior Lecturer,
South Ural State University,
Chelyabinsk, Russia
E-mail: ufimtcevaov@susu.ru

Subject of research: an array of data including socio-economic parameters of transactions in the primary housing market of the Chelyabinsk region.

Purpose of research: to create a neural network model for predicting real estate sales volumes with a time horizon of one year based on retrospective data. The predictive model should be easy to set up and operate, as well as provide for an intuitive user interface.

Research methods: preliminary data processing in accordance with the CRISP-DM standard, correlation analysis, normalization, k-fold cross-validation, construction of a fully connected neural network with two hidden layers and a sigmoidal activation function, the L-BFGS optimization method.

Objects of research: forecasting sales volumes in the regional primary housing construction market.

Research findings: a neural network forecasting model has been created using the Loginom Community analytical platform, which has a modular structure, which ensures its versatility. The model was successfully tested on a validation sample, demonstrating a prediction accuracy of 88.33%.

Keywords: forecasting, perceptrons, data analysis, decision making, time series, neural network learning algorithms.

ВВЕДЕНИЕ

Актуальность настоящего исследования обусловлена необходимостью повышения качества финансового планирования в строительных организациях в контексте введения федерального закона № 214, регулирующего участие в долевом строительстве. В соответствии с указанным законом при оплате по договору долевого участия должны использоваться счета эскроу. Для строительства жилья компания берет банковский кредит, процентная ставка по которому зависит от уровня наложения эскроу-счетов.

В процессе планирования, при разработке финансовой модели и модели движения денежных средств для расчета процентов по кредитным обязательствам и сопутствующих издержек необходимо учитывать объем средств на эскроу-счетах, который напрямую коррелирует с объемом реализованной жилой недвижимости.

В настоящее время региональные девелоперские компании осуществляют прогнозирование объема продаж на основании

показателей, достигнутых в текущем году, используя субъективные экспертные оценки. Данный метод прогнозирования требует высокого уровня квалификации специалистов и не имеет формализованной структуры. Вследствие этого планируемые финансовые показатели характеризуются значительной степенью неопределенности и являются преимущественно ориентировочными.

В связи с этим возникает потребность в разработке и внедрении более точных и формализованных методик прогнозирования объемов реализации недвижимости, основанных на комплексном анализе ретроспективных данных об операциях купли-продажи.

Согласно результатам анализа научной литературы [2–8], в практике прогнозирования применяются статистические и структурные модели.

Статистические модели осуществляют анализ взаимосвязей между зависимой переменной и определяющими факторами посредством использования математических



зависимостей, базирующихся на двух ключевых предпосылках: во-первых, предполагается определенное распределение данных, во-вторых, выдвигается гипотеза о характере взаимосвязи между переменными. Нарушение данных предпосылок может привести к некорректным результатам прогнозирования.

Многие реальные процессы обладают нелинейными и хаотическими характеристиками, что существенно усложняет их анализ с применением традиционных стохастических методов. В таких ситуациях более целесообразным представляется использование структурных моделей, основанных на методах машинного обучения, таких как нейронные сети.

При разработке модели прогнозирования объема продаж недвижимости в новостройках на уровне отдельной региональной компании были учтены следующие аспекты:

1. Для создания модели прогнозирования объема продаж региональная компания должна применять информационный подход, основанный на сборе и обработке данных из различных источников.

2. Модель, разработанная на основе методов машинного обучения, позволяет учитывать сложные взаимосвязи и зависимости между множественными факторами, влияющими на объем продаж.

3. Нейронные сети обладают адаптивными свойствами, что позволяет им автоматически выявлять скрытые закономерности в исходных данных. Эти сети демонстрируют высокую эффективность при анализе больших объемов информации и выявлении сложных нелинейных зависимостей.

Для реализации поставленной цели была задействована аналитическая платформа Loginom, обладающая визуальным конструктором для настройки процессов подготовки данных и моделирования. Данная функция способствует значительному сокращению временного интервала между тестированием гипотез и внедрением актуальных бизнес-процессов. В качестве инструмента для прогнозирования был использован многослойный персептрон (MLP), интегрированный в аналитическую платформу.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Подготовка данных

Гипотеза исследования формулируется следующим образом: динамика объема продаж недвижимости определяется совокупностью факторов, включающих макроэкономические индикаторы федерального уровня, региональные макроэкономические

параметры, корпоративные факторы, обусловленные спецификой деятельности конкретной организации, а также индивидуальные предпочтения потенциальных покупателей.

К внешним макроэкономическим факторам, регулируемым на государственном уровне, относятся процентная ставка Центрального банка и уровень инфляции. Официальная статистическая информация по этим факторам доступна на официальном веб-сайте Единой межведомственной информационно-статистической системы (ЕМИСС) [1].

Региональные макроэкономические факторы, влияющие на финансовые возможности покупателей, включают среднедушевые денежные доходы, размер материнского капитала, использованного для приобретения жилья, и способы оплаты, которые могут быть представлены в виде ипотечного кредитования или единовременной оплаты стоимости квартиры.

Корпоративные факторы включают стоимость квадратного метра недвижимости и общую стоимость квартиры.

Индивидуальные предпочтения потенциальных покупателей оцениваются косвенно на основе следующих показателей: комплексная оценка объектов недвижимости, которая включает местоположение и статус дома, количество комнат в квартире, а также проведение акций для стимулирования покупки с конкретным клиентом.

База данных включала свыше 5000 записей о ежедневных операциях по продаже недвижимости в период с марта 2018 по апрель 2025 года. Прогностическая модель была разработана на основе стандартизированной методологии анализа данных CRISP-DM (Cross-Industry Standard Process for Data Mining). В процессе подготовки данных были устранены аномальные значения, заполнены пропуски, осуществлена кодификация категориальных признаков и произведена группировка данных по временным интервалам – году и месяцу.

Корреляционный анализ выявил, что на объем продаж недвижимости оказывают значительное влияние следующие факторы:

- количество и размер ипотечных кредитов;
- средняя стоимость квадратного метра;
- проведение маркетинговых кампаний;
- процентная ставка Центрального банка;
- уровень инфляции;
- среднедушевые денежные доходы.

Обучающая выборка состояла из данных за период с марта 2018 по декабрь 2024 года, которые были случайным образом разделены на обучающую и тестовую подвыборки в

соотношении 90 % к 10 % соответственно. Данные за период с января по апрель 2025 года были использованы в качестве валидационной выборки.

Нейросетевая модель прогнозирования

На рисунке 1 приведена обобщенная структура предлагаемой модели прогнозирования.

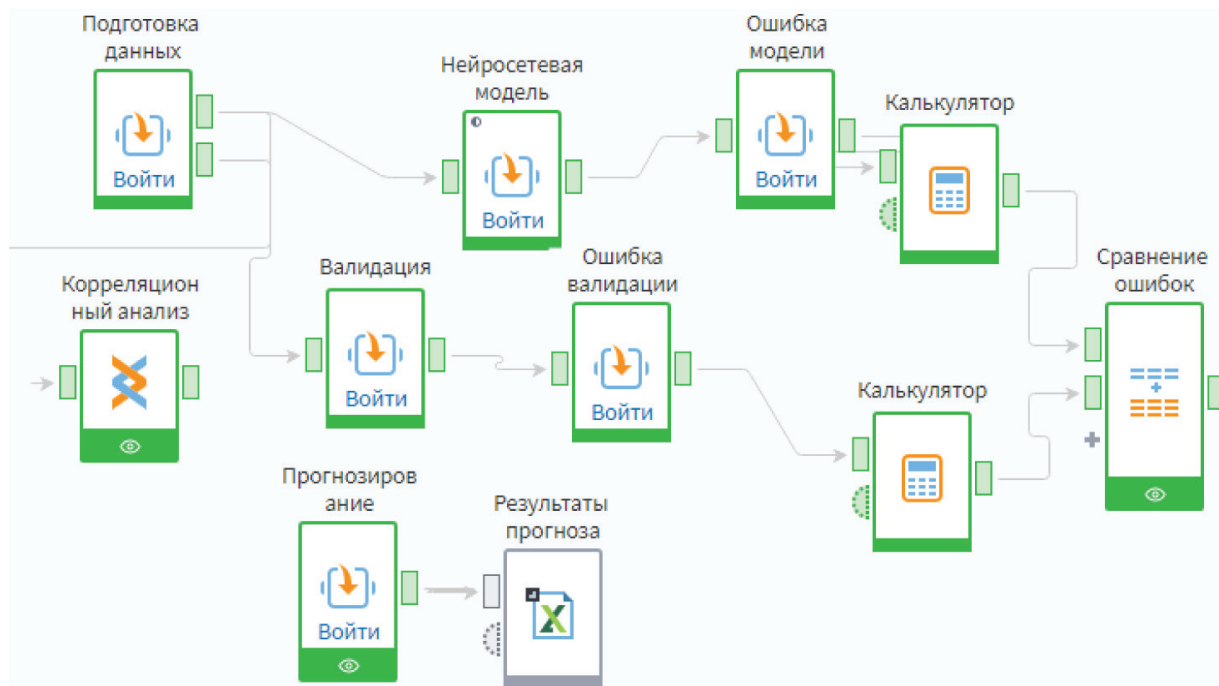


Рисунок 1. Обобщенная структура модели прогнозирования

Прогностическая модель имеет модульную архитектуру и функционирует на базе аналитической платформы Loginom Community.

Основные функции модели:

- сбор и подготовка данных для проведения прогнозного анализа;
- обучение нейросетевой модели;
- оценка и уточнение модели;
- выполнение прогнозных расчетов.

Модульная структура обеспечивает универсальность модели посредством выделения отдельных независимых модулей в рамках общей архитектуры. Архитектурная реализация модели базируется на многослойном персептроне с активационной функцией сигмоидального типа.

Обучение нейронной сети осуществлялось на основе исторических данных с применением алгоритма L-BFGS [9]. Последующая валидация модели проводилась на новых примерах для оценки её способности к обобщению. Во избежание переобучения был использован алгоритм k-fold кросс-валидации с пятикратной повторной выборкой.

Многослойный персептрон (рисунок 2) с единственным нейроном на выходе описывается в следующем компактном виде (1)

$$F(X, W) = \varphi(\sum_k w_{ok} \varphi(\sum_j w_{kj} \varphi(\dots \varphi(\sum_i w_{li} x_i))), \quad (1)$$

где $\varphi(\cdot)$ – сигмоидальная функция активации; w_{ok} – синаптический вес связи между нейроном k последнего скрытого слоя и единственным выходным нейроном o и т. д. для всех остальных синаптических весов; x_i – i -й элемент входного вектора X . Матрица W содержит полное множество синаптических весов, упорядоченных сначала по слоям, затем по нейронам каждого отдельного слоя и, наконец, по синапсам отдельных нейронов [9].

Не существует строго установленных критериев для определения оптимального количества нейронов в нейронной сети. Ключевое значение имеют веса матрицы W , число которых определяется количеством связей между нейронами, а не общим числом нейронов. В соответствии с источником [9], на практике для хорошего обобщения достаточно, чтобы размер обучающего множества n удовлетворял соотношению (2):

$$n = O(W/\varepsilon), \quad (2)$$

где W – общее количество свободных параметров (т. е. синаптических весов и порогов); ε – допустимая точность ошибки; $O(\cdot)$ – порядок заключенной в скобки величины.

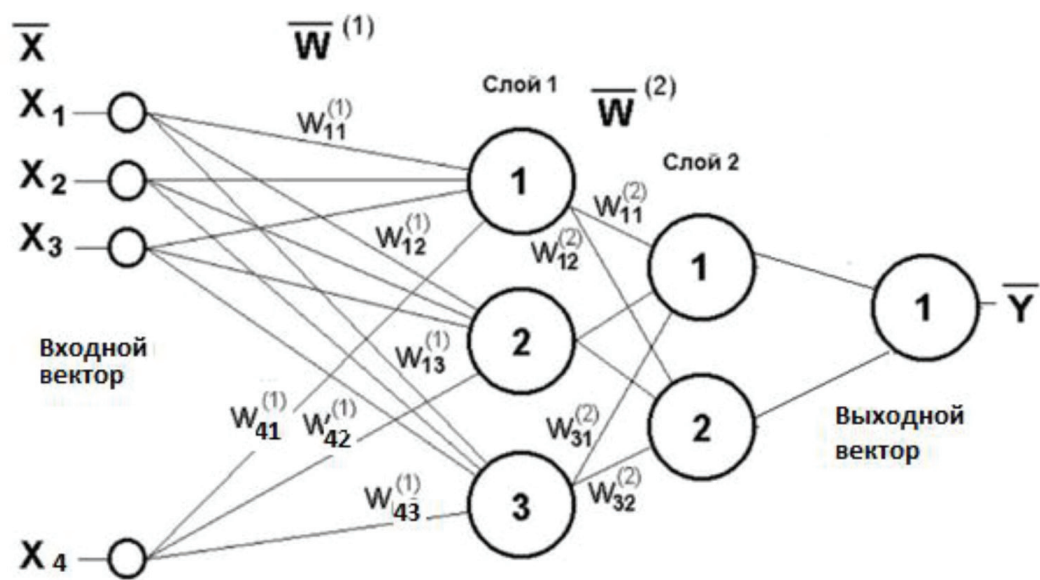


Рисунок 2. Архитектура нейронной сети

Для двухслойной нейронной сети свободные параметры могут быть распределены различными способами. Конкретная конфигурация определяется в процессе обучения. При настройке весовых коэффициентов матрицы (W) необходимо рассмотреть несколько возможных вариантов.

Для выбора оптимальной архитектуры сети был проведен ряд экспериментов с различными конфигурациями входных данных, количеством слоев и числом нейронов, примеры которых приведены в таблице 1.

Параметры сети нормализованы. Метод нормализации – стандартизация.

Таблица 1. Варианты конфигураций нейронных сетей

Входные факторы	Архитектура НС	Средняя относительная ошибка		Ошибка MRPE, %		Значимость параметров НС	Примечание
		обучения	обобщения	модели	валидации		
Номер периода, средняя цена кв. м, оплата наличными, кредиты, ставка ЦБ, % инфляции, среднедушевые доходы	7×10×9×1	0,18	0,22	-	-	Незначимы	НС переобучена
Номер периода, сумма кредитов, средняя цена кв. м, ставка ЦБ, количество акций, % инфляции	6×9×10×1	0,42	1,21	-	-	Значимы	НС переобучена
Номер периода, сумма кредитов, средняя цена кв. м, ставка ЦБ, количество акций	5×11×1	0,38	0,19	7,73	6,46	Низкая	НС обучена
Номер периода, сумма кредитов, средняя цена кв. м, ставка ЦБ, комплекс	5×17×8×1	1,49	1,14	12,6	11,71	Значимы	НС обучена
Номер периода, сумма кредитов, средняя цена кв. м, ставка ЦБ	4×11×14×1	0,63	0,053	11,91	11,57	Значимы	НС обучена

Стандартизация данных преобразует их к формату с нулевым средним и стандартным отклонением, равным 1, независимо от распределений и единиц измерения. Это упрощает многомерный анализ. Формула стандартизации:

$$z_i = \frac{x_i - X_{\text{среднее}}}{\sigma_x}, \quad (3)$$

где x_i – исходное значение признака; $X_{\text{среднее}}$ – среднее значение признака по всему набору данных; σ_x – стандартное отклонение признака по всему набору данных.

В процессе обучения формируется обобщающая способность сети, но возможны недообучение и переобучение. Недообучение приводит к плохой работе на обучающих данных, переобучение – на тестовых. Для предотвращения переобучения использовалась k-fold кросс-валидация с количеством фолдов, равным 5.

Для выявления статистически значимых различий между группами данных использовался G-критерий хи-квадрат, метод особенно эффективен при наличии аномалий в

выборке. В качестве меры вероятности ошибки используется р-значение, которое обратно пропорционально значимости полученного результата: чем меньше значение р, тем выше статистическая значимость выявленного различия.

На основе сравнительного анализа нескольких вариантов моделей (см. таблицу 1) была выбрана модель с архитектурой $4 \times 11 \times 14 \times 1$, которая показала хорошие результаты по всем метрикам качества: MRPE (средняя относительная процентная ошибка) составила около 11,57 % на валидационных данных, взвешенная абсолютная процентная ошибка WAPE=5,66 % и среднеквадратичная процентная ошибка MSPE=1,40 % также указали на высокую точность модели (рисунок 3).

Проведенная оценка ошибок подтвердила устойчивость модели к переобучению, а также ее применимость в условиях наличия шума и аномалий в данных. При этом особое значение имело то, что ошибка на тестовой выборке оказалась ниже, чем на обучающей, что свидетельствует о хорошем балансе между сложностью модели и её способностью адаптироваться к новым данным.

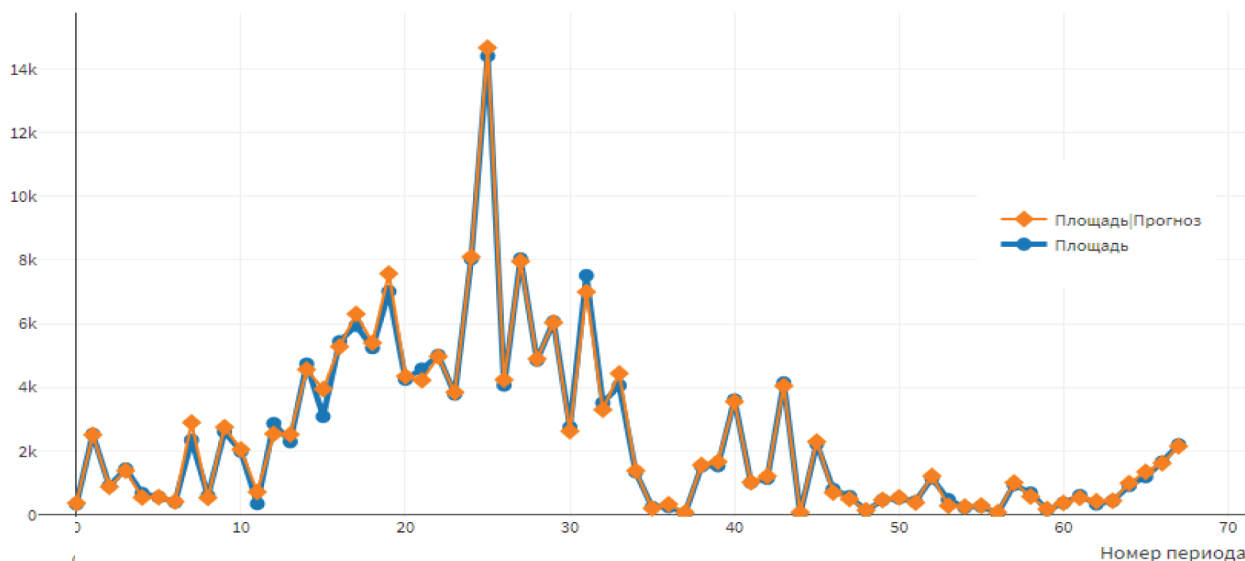


Рисунок 3. Сравнение реальной и спрогнозированной площадей

ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

При принятии управленческих решений в любой организации важно использовать удобный и надежный инструмент прогнозирования, обеспечивающий высокую степень достоверности результатов.

Разработанная модель отличается универсальностью и может использоваться не только в текущих экономических условиях, но и при изменении ключевых макроэкономических факторов. Это делает её применимой к

различным строительным проектам, реализуемым в Челябинской области, а также позволяет адаптировать ее под различные сценарии развития рынка недвижимости: оптимистичный, реалистичный и пессимистичный.

Следует подчеркнуть, что качество прогноза, получаемого посредством использования нейронной сети, в значительной степени определяется полнотой и адекватностью исходных данных. Результаты прогноза также зависят от внешних экономических факторов,

что имеет существенное практическое значение для строительной организации.

СПИСОК ЛИТЕРАТУРЫ

1. ЕМИСС : государственная статистика. – URL: <https://www.fedstat.ru> (дата обращения: 18.04.2025).
2. Зайнакова, С. Р. Рынок жилья в России: современное состояние, проблемы и прогнозы / С. Р. Зайнакова // Социально-экономические науки и гуманитарные исследования. – 2014. – № 1. – С. 58–62.
3. Зинич, Л. В. Анализ состояния и тенденций рынка жилой недвижимости в России: вызовы и возможности / Л. В. Зинич, Е. С. Петров // Жилищные стратегии. – 2025. – Т. 12, № 1. – С. 39–58.
4. Свиридов, А. В. Прогнозирование развития локального рынка жилой недвижимости : диссертация на соискание ученой степени кандидата экономических наук : 05.13.18 / А. В. Свиридов. – Москва. – 2018. – 189 с.
5. Стерник, Г. М. Анализ рынка недвижимости для профессионалов / Г. М. Стерник, С. Г. Стерник. – Москва : Экономика, 2009. – 606 с.
6. Стерник, Г. М. Методика прогнозирования цен на жилье в зависимости от типа рынка / Г. М. Стерник // Имущественные отношения в РФ. – 2010. – № 12. – С. 43–47.
7. Стерник, Г. М. Статистический подход к прогнозированию цен на жилье / Г. М. Стерник // Экономика и математические методы. – 1998. – № 1. – С. 85–90.
8. Стерник, Г. М. Методология прогнозирования российского рынка недвижимости. Ч. 1. Основные допущения, ограничения и рабочие гипотезы / Г. М. Стерник // Механизация строительства. – 2013. – № 8. – С. 53–63.
9. Хайкин, С. Нейронные сети : полный курс : перевод с английского / С. Хайкин. – 2-е издание. – Москва : Вильямс, 2006. – 1104 с.