

Министерство образования Республики Беларусь
Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

УДК 004.383

Голубко
Дмитрий Владимирович

Система менеджмента недвижимости и анализа цен на рынке

ДИССЕРТАЦИЯ
на соискание академической степени
магистра технических наук
по специальности 1-40 80 05 — Программная инженерия

Научный руководитель

Смолякова О. Г.
к.т.н., доцент

Минск 2020

СОДЕРЖАНИЕ

| | |
|---|----|
| ПЕРЕЧЕНЬ УСЛОВНЫХ ОБОЗНАЧЕНИЙ | 3 |
| ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ | 4 |
| ВВЕДЕНИЕ | 5 |
| 1 Анализ существующих алгоритмов | 7 |
| 1.1 Регрессионный анализ | 7 |
| 1.2 Нейросетевой анализ | 15 |
| 2 Экспериментальная часть | 16 |
| 2.1 Подготовка данных | 16 |
| 2.2 Регрессионный анализ | 17 |
| 2.3 Анализ с использованием нейронных сетей | 17 |
| 2.4 Сравнение результатов | 20 |
| ЗАКЛЮЧЕНИЕ | 22 |
| СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ | 23 |

ПЕРЕЧЕНЬ УСЛОВНЫХ ОБОЗНАЧЕНИЙ

МНК – Метод наименьших квадратов

ПС – программное средство

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Цель диссертационной работы – провести анализ применяемых моделей и алгоритмов, которые используются при прогнозировании цен на рынке недвижимости, проверить адекватность применения методов эконометрического анализа для оценки объектов недвижимости и построение на их основе модели стоимости, применить существующие алгоритмы на подготовленной выборке данных. На основе проведенного анализа выполнить сравнение алгоритмов, создать соответствующее ПС для управления объектами недвижимости с возможностью предварительной оценки стоимости недвижимости. Результаты исследования могут быть полезны для прогнозирования ценообразования на рынке недвижимости, а также при оценке стоимости объектов недвижимости.

Для достижения поставленной цели необходимо решить следующие задачи:

- а) Провести анализ применяемых моделей и алгоритмов, которые используются при прогнозировании цен на рынке недвижимости.
- б) Реализовать предложенные модели и провести экспериментальные исследования на основе накопленных данных.
- в) На основе сравнительного анализа выбрать более эффективный и достоверный.
- г) Реализовать соответствующее ПС для управления и предварительной оценки стоимости недвижимости.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников, списка публикаций автора и приложений. В первой главе представлен анализ существующих применяемых моделей и алгоритмов. Вторая глава посвящена проведению экспериментального исследования, реализацией выбранных алгоритмов анализа, проведению их сравнительного анализа. В третьей главе предложена разработка программной реализации системы управления недвижимостью с возможностью предварительной оценки стоимости объектов недвижимости.

ВВЕДЕНИЕ

Переход к рыночным отношениям в экономике и научно-технический прогресс чрезвычайно ускорили темпы внедрения во все сферы социально-экономической жизни общества последних научных разработок в области информационных технологий.

Рынок недвижимости представляет собой механизм, обслуживающий и регулирующий отношения по купле, продаже и аренде недвижимости на основе спроса и предложения. В мировой практике можно выделить следующие типы рынков недвижимости:

- рынок жилой недвижимости;
- рынок коммерческой недвижимости, приносящей доход ее владельцу (офисные, торговые, производственные, складские помещения);
- рынок земельных участков.

Рынок недвижимости делится на первичный и вторичный. Объектом сделок на первичном рынке является новая недвижимость, т.е. только что построенные дома, квартиры, офисные и другие помещения. Их могут продавать застройщики, инвесторы, финансировавшие строительство. На вторичном рынке предоставлено жилье и помещения, которыми уже пользовались по основному назначению. Первичный рынок отражает объемы созданной жилой недвижимости, а объем вторичного рынка определяется другими факторами:

- изменением благосостояния населения;
- доходностью различных инвестиционных объектов;
- мобильностью трудовых ресурсов;
- событиями человеческой жизни (свадьба, развод, рождение ребенка в семье, смена места жительства и др.).

Объекты недвижимости занимают значительную часть ресурсов экономики любой страны. Оценка стоимости – длительный и сложный процесс установления денежного эквивалента стоимости объекта недвижимости. Она требует высокой квалификации оценщика, владеющего методами и инструментарием оценочной деятельности, знающего состояние рынка недвижимости и особенно нужного сегмента, детального значения правовых особенностей сделок с недвижимостью и др [1]. Практика показывает, что для оценки стоимости объекта недвижимости, специалисту требуется значительное время.

С развитием теоретических подходов для создания адекватных моделей поведения рынка недвижимости в западных странах и США одновременно происходило активное внедрение новых интеллектуальных компьютерных

технологий в практику принятия финансовых и инвестиционных решений. Вначале в виде экспертных систем и баз знаний, а затем с конца 80-х - нейросетевых технологий, которые являются адекватным аппаратом для решения задач прогнозирования.

Начало исследования методов обработки информации, называемых сегодня нейросетевыми, было положено несколько десятилетий назад. С течением времени интерес к нейросетевым технологиям то ослабевал, то вновь возрождался. Такое непостоянство напрямую связано с практическими результатами проводимых исследований.

Цена на жилье - вопрос крайне сложный, изучение основных факторы влияния и выяснение правил изменения имеют важное теоретическое и практическое значение для способствования устойчивому и здоровому развитию рынка жилой недвижимости. Автоматизация позволит ускорить процесс принятия решения, учесть большее количество факторов и снизить уровень субъективности.

1 АНАЛИЗ СУЩЕСТВУЮЩИХ АЛГОРИТМОВ

1.1 Регрессионный анализ

В статистическом моделировании регрессионный анализ – это набор статистических процедур для изучения зависимостей между случайными переменными. Он включает в себя множество методов моделирования и анализа взаимосвязей между зависимой переменной и одной или несколькими независимыми переменными, называемых также предикторами или регрессорами.

Регрессионный анализ помогает понять, как значение зависимой переменной изменяется при изменении одной из независимых переменных, в то время как другие независимые переменные остаются фиксированными.

Чаще всего в регрессионном анализе оценивается условное математическое ожидание зависимой переменной с учетом значений, принимаемых независимыми переменными. Во всех случаях оценивается функция математического ожидания зависимой переменной от независимых переменных, называемая функцией регрессии.

Регрессионный анализ широко используется для численного предсказания, классификации и прогнозирования, где его применение существенно перекрывается с областью машинного обучения.

В настоящее время разработано много методов регрессионного анализа. Наиболее популярными из них являются простая и множественная линейная регрессия, среднеквадратическая и логистическая регрессия. Эти модели, являются параметрическими в том смысле, что функция регрессии определяется конечным числом неизвестных параметров, которые оцениваются на основе данных.

В математической статистике линейная регрессия представляет собой метод аппроксимации зависимостей между входными и выходными переменными на основе линейной модели. Является частью более широкой статистической методики, называемой регрессионным анализом.

В регрессионном анализе входные (независимые) переменные называются также предикторными переменными или регрессорами, а зависимые переменные — критериальными.

Если рассматривается зависимость между одной входной и одной выходной переменными, то имеет место простая линейная регрессия. Для этого определяется уравнение регрессии и строится соответствующая прямая, известная как линия регрессии(рис. 1.1).

$$y = ax + b \tag{1.1}$$

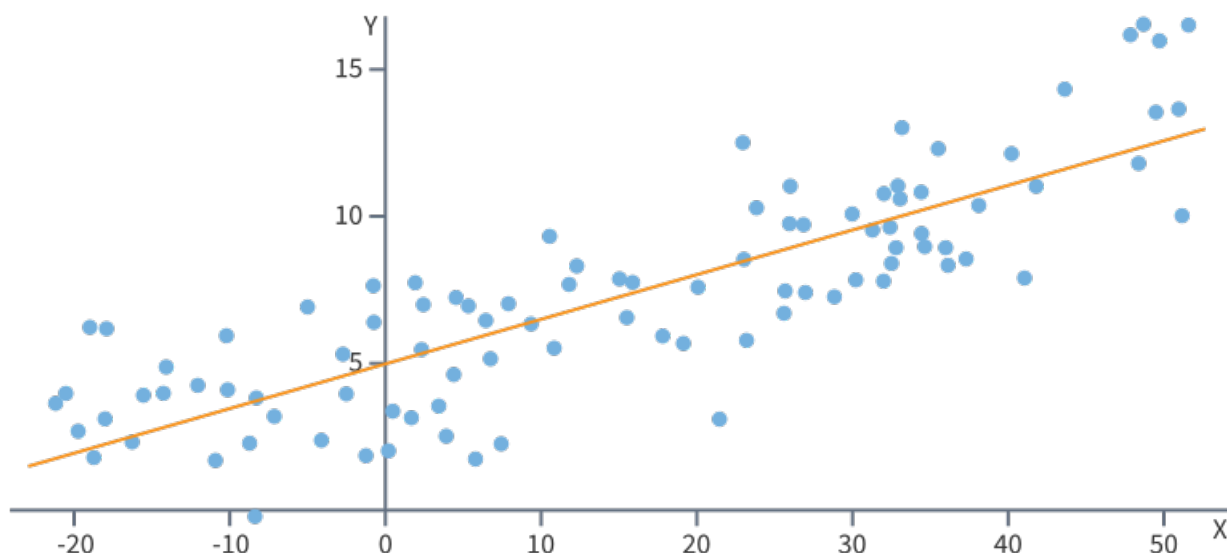


Рисунок 1.1 – Линия регрессии

Коэффициенты a и b , называемые также параметрами модели, определяются таким образом, чтобы сумма квадратов отклонений точек, соответствующих реальным наблюдениям данных, от линии регрессии была бы минимальной. Коэффициенты обычно оцениваются методом наименьших квадратов.

Метод наименьших квадратов (МНК) — математический подход для оценки параметров моделей (например, регрессионной) на основании экспериментальных данных, содержащих случайные ошибки.

Если данные известны с некоторой погрешностью, то вместо неизвестного точного значения параметра модели используется приближенное. Поэтому параметры модели должны быть рассчитаны так, чтобы минимизировать разницу между экспериментальными данными и теоретическими (вычисленными при помощи предложенной модели).

Мерой рассогласования между фактическими значениями и значениями, оцененными моделью в методе наименьших квадратов, служит сумма квадратов разностей между ними, т.е.:

$$\sum_{i=1}^N (y' - y)^2 \quad (1.2)$$

где y' — оценка, полученная с помощью модели;
 y — фактическое наблюдаемое значение.

Очевидно, что лучшей будет та модель, которая минимизирует данную сумму.

Важнейшим применением МНК в анализе данных является линейная регрессия, где параметры регрессионной модели вычисляются таким образом, чтобы сумма квадратов расстояний от линии регрессии до фактических значений данных была минимальной [2].

Если ищется зависимость между несколькими входными и одной выходной переменными, то имеет место множественная линейная регрессия. Соответствующее уравнение имеет вид:

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (1.3)$$

где n — число входных переменных.

Очевидно, что в данном случае модель будет описываться не прямой, а гиперплоскостью. Коэффициенты уравнения множественной линейной регрессии подбираются так, чтобы минимизировать сумму квадратов отклонения реальных точек данных от этой гиперплоскости.

Преимущество множественной линейной регрессии по сравнению с простой заключается в том, что использование в модели нескольких входных переменных позволяет увеличить долю объяснённой дисперсии выходной переменной, и таким образом улучшить соответствие модели данным. Т.е. при добавлении в модель каждой новой переменной коэффициент детерминации растёт.

Коэффициент детерминации – Статистический показатель, отражающий объясняющую способность регрессии $a: X \rightarrow Y$ и равный отношению суммы квадратов регрессии SSR к общей вариации SST :

$$r^2 = \frac{SSR}{SST} = \frac{\sum_i (a(x_i) - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \quad (1.4)$$

где $x_l = (x_i, y_i)_{i=1}^l$ — набор данных из l наблюдений.

x_i — вектор признаков i -го наблюдения,

$y_i \in Y$ — y_i принадлежит Y ;

Данный показатель является статистической мерой согласия, с помощью которой можно определить, насколько уравнение регрессии соответствует реальным данным.

Коэффициент детерминации изменяется в диапазоне от 0 до 1. Если он равен 0, это означает, что связь между переменными регрессионной мо-

дели отсутствует и вместо нее для оценки значения выходной переменной можно использовать простое среднее ее наблюдаемых значений. Напротив, если коэффициент детерминации равен 1, это соответствует идеальной модели, когда все точки наблюдений лежат точно на линии регрессии, т.е. сумма квадратов их отклонений равна 0.

На практике, если коэффициент детерминации близок к 1, это указывает на то, что модель работает очень хорошо (имеет высокую значимость), а если к 0, то это означает низкую значимость модели, когда входная переменная плохо «объясняет» поведение выходной, т.е. линейная зависимость между ними отсутствует. Очевидно, что такая модель будет иметь низкую эффективность.

В некоторых случаях коэффициент детерминации может принимать небольшие отрицательные значения, если модель получилась «бесполезной» и ее предсказания хуже, чем оценки на основе среднего значения [3].

Линейная регрессия была первым видом регрессионного анализа, который был тщательно изучен и начал широко использоваться в практических приложениях. Это связано с тем, что в линейных моделях оценивание параметров проще, а также с тем, что статистические свойства полученных оценок легче определить.

Линейная регрессия имеет много практических применений. Большинство приложений попадают в одну из двух широких категорий:

- Если целью является прогнозирование, линейную регрессию можно использовать для подгонки модели к наблюдаемому набору данных.

- Если цель заключается в том, чтобы объяснить изменчивость выходной переменной, можно применить линейный регрессионный анализ для количественной оценки силы взаимосвязи между выходной и входными переменными.

Среднеквадратическая регрессия – разновидность регрессии, где при определении параметров модели используется обобщение метода наименьших квадратов (МНК) – метод наименьших средних квадратов.

Иными словами, в процессе подгонки модели к данным минимизируется не сумма квадратов остатков регрессии, а их средний квадрат:

$$E[(Y - F(X))^2] \quad (1.5)$$

где E — операция усреднения,

Y — зависимая переменная,

X — вектор независимых переменных.

Это становится возможным благодаря тому, что МНК допускает широкое обобщение, когда вместо минимизации суммы квадратов остатков можно минимизировать их некоторую положительно определённую квадратичную форму.

Смысл данного подхода заключается в том, что к результатам классического МНК (т.е. квадратам остатков) применяется дополнительное линейное преобразование - усреднение. Как известно, одним из предположений регрессии является предположение о нормальности остатков, которое в практических случаях не соблюдается. Усреднение позволяет снизить степень влияния отклонения остатков от нормального распределения на качество построенной модели.

Метод наименьших средних квадратов был сформулирован Бернардом Уидроу и Тедом Хоффом в 1960 году и применён при обучении нейронных сетей с помощью алгоритма обратного распространения ошибки.

В математическом и статистическом моделировании зависимой (выходной) называется переменная модели, которая зависит от входных переменных и случайных факторов, воздействующих на моделируемый процесс или объект.

Выходная переменная представляет результаты работы модели. Она изменяется (варьирует) под воздействием изменения входной переменной и случайных факторов. Изучение изменчивости выходной переменной при изменении входной и является целью моделирования.

В статистическом моделировании и машинном обучении независимой (входной) переменной называют величину, от которой зависит изменение выходной переменной, при этом целью построения модели является аппроксимация этой зависимости.

Аппроксимация – математический метод, в основе которого лежит замена одних математических объектов другими, близкими к исходным в том или ином смысле, но более простыми. [4]

Аппроксимация позволяет исследовать числовые характеристики и качественные свойства объекта, сводя задачу к изучению более простых или более удобных объектов (например, тех, параметры которых легко вычисляются или известны заранее).

Например, в линейной регрессии некоторая неизвестная функция, описывающая реальные наблюдения, аппроксимируется уравнением прямой, а если наблюдаемые данные носят нелинейный характер, то полиномами и т.д.

моделью, которая использует логистическую функцию для моделирования зависимости выходной переменной от набора входных в случае, когда первая является бинарной.

Это разновидность множественной регрессии, общее назначение которой состоит в анализе связи между несколькими независимыми переменными (называемыми также регрессорами или предикторами) и зависимой переменной. Регрессия в общем виде применяется, когда входные и выходная переменные непрерывные. А логистическая регрессия лучшим образом подходит, когда выходная переменная принимает только два значения.

Важность логистической регрессии обусловлена тем, что многие задачи анализа данных могут быть решены с помощью бинарной классификации или сведены к ней.

Например, с помощью логистической регрессии можно оценивать вероятность наступления (или не наступления) некоторого события: пациент болен (здоров), заемщик вернул кредит (допустил просрочку) и т.д. Благодаря этому логистическую регрессию можно рассматривать как мощный инструмент поддержки принятия решений.

Как известно, все регрессионные модели могут быть записаны в виде формулы:

$$y = F(x_1, x_2, \dots, x_n) \quad (1.6)$$

Например, если рассматривается исход по займу, задается переменная y со значениями 1 и 0, где 1 означает, что соответствующий заемщик расплатился по кредиту, а 0 — что имел место дефолт.

Однако здесь возникает проблема: множественная регрессия не «знает», что переменная отклика бинарная по своей природе. Это неизбежно приведет к модели с предсказываемыми значениями большими 1 и меньшими 0. Но такие значения вообще не допустимы для первоначальной задачи. Таким образом, множественная регрессия просто игнорирует ограничения на диапазон значений для y .

Для решения проблемы задача регрессии может быть сформулирована иначе: вместо предсказания бинарной переменной мы предсказываем непрерывную переменную со значениями на отрезке $[0,1]$ при любых значениях независимых переменных. Это достигается применением следующего регрессионного уравнения (логит-преобразование):

$$p = \frac{1}{1 + e^{-y}} \quad (1.7)$$

где p — вероятность того, что произойдет интересное событие,
 e — основание натуральных логарифмов 2,71...,
 y — стандартное уравнение регрессии.

Зависимость, связывающая вероятность события и величину y , показана на следующем графике:

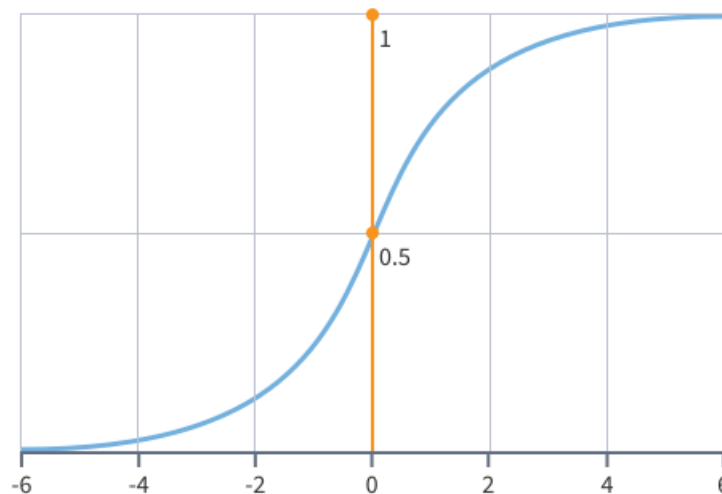


Рисунок 1.2 – Зависимость, связывающая вероятность события и величину y

Преобразование вида:

$$P' = \log_e \left(\frac{P}{1 - P} \right) \quad (1.8)$$

называют логистическим, или логит-преобразованием.

Существует несколько способов нахождения коэффициентов логистической регрессии. На практике часто используют метод максимального правдоподобия. Он применяется в статистике для получения оценок параметров генеральной совокупности по выборочным данным.

Логистическая регрессия является традиционным статистическим инструментом для расчета коэффициентов (баллов) скоринговой карты на основе накопленной кредитной истории.

Генеральная совокупность — это совокупность всех объектов или наблюдений, относительно которых исследователь намерен делать выводы при решении конкретной задачи. В ее состав включаются все объекты, которые подлежат изучению.

Объем генеральной совокупности может быть очень велик, и на практике рассмотреть все ее элементы не представляется возможным. Поэтому

обычно из генеральной совокупности извлекаются выборки, на основе анализа которых аналитик пытается сделать вывод о свойствах всей совокупности, скрытых в ней закономерностях, действующих правилах и т.д. При этом выборки должны быть репрезентативными.

Регрессионный анализ является одним из наиболее распространенных методов обработки результатов экспериментов при изучении зависимостей в естественных науках, экономике, технике и др. областях.

В аналитических технологиях Data Mining элементы регрессионного анализа широко используются для решения задач прогнозирования, оценивания, классификации, выявления зависимостей между признаками.

Основы регрессионного анализа были заложены А. Лежандром и Карлом Гауссом в их работах по методу наименьших квадратов в начале 19 в. [5]

В целях оценки недвижимости может применяться либо многофакторный, либо однофакторный регрессионный анализ. В первом случае строится множественная регрессионная модель, описывающая зависимость стоимости оцениваемого объекта от нескольких независимых определяющих факторов, значения которых определяются из анализа рыночных данных. Этими факторами могут быть как физические характеристики объекта (площадь, качество отделки и т.п.), так и характеристики его местоположения (удаленность от транспортных магистралей, экологическая обстановка и т.п.).

При однофакторном регрессионном анализе рассматривается зависимость переменной – стоимости единицы сравнения – от одной независимой (контролируемой) переменной. Значения остальных независимых переменных считаются фиксированными. В качестве независимой переменной X обычно используется показатель «общая площадь», за зависимую переменную Y принимается показатель «стоимость 1м кв. общей площади».

В случае, когда рассматривается зависимость между одной зависимой переменной Y и несколькими независимыми переменными X_1, X_2, \dots, X_n говорят о множественной линейной регрессии.

Исходя из проведенного анализа способов применения различных методов регрессии можно сделать вывод, что наиболее подходящим способом для анализа процесса формирования цен на недвижимость является множественный линейный регрессионный анализ.

1.2 Нейросетевой анализ

2 ЭКСПЕРЕМЕНТАЛЬНАЯ ЧАСТЬ

2.1 Подготовка данных

В связи с тем, что рынок вторичного жилья лучше соответствует рыночным принципам формирования цен на основе спроса и предложения, в отличие от цен, устанавливаемых компанией-застройщиком жилья на первичном рынке, определение рыночной стоимости, как наиболее вероятной цены продажи объекта недвижимости, более целесообразно провести на примере объектов недвижимости вторичного рынка жилья. При создании модели оценки жилой недвижимости в качестве входных параметров были включены факторы, представленные в таблице 2.1:

Таблица 2.1 – Назначение папок, используемых в проекте

| Параметр | Описание |
|-----------------------|---|
| City | Город |
| State | Штат |
| Building area | Общая площадь, кв.м |
| Carport count | Количество парковочных мест под навесом |
| Garage count | Количество гаражей |
| Open spaces count | Количество открытых парковочных мест |
| Property Type | Тип недвижимости(дом, таунхаус, апартаменты, итд.) |
| Bathrooms | Количество санузлов |
| Bedrooms | Количество спален |
| Number of sales | Количество проданных жилых объектов в данном городе за 5 лет |
| Average Price | Средняя цена проданных жилых объектов в данном городе за 5 лет |
| Median Rental Price | Медианная стоимость аренды недвижимости в данном городе за 5 лет |
| Change in Rental Rate | Процент изменения средней арендной платы в данном городе за 5 лет |
| Sold Price | Стоимость данного объекта |

Как можно заметить, полученные данные являются как количественными, так и качественными. Количественные данные остаются без изменений, для качественных(город, штат, тип недвижимости) были введены числовые характеристики. Исходные данные были взяты из базы проданной недвижимости в Австралии за период с 2017 по 2020 год. Всего было собрано дан-

ных о более чем 6 тысяч проданных и продающихся объектов недвижимости. Данные хранятся в виде таблицы в формате, часть которой показана на рисунке 2.1

2.2 Регрессионный анализ

В качестве первого способа анализа рынка недвижимости было решено использовать регрессионный анализ.

Виды регрессионного анализа,
почему выбран именно такой,
сравнение разных способов на основе других статей(под вопросом?)

Регрессионный анализ представляет собой процесс подбора математического выражения для функции нескольких независимых переменных. Основная задача регрессионного анализа заключается в исследовании влияния нескольких независимых переменных X на зависимую переменную Y .

Используемые технологии

Результаты

2.3 Анализ с использованием нейронных сетей

Для сравнения также было решено разработать методику оценки стоимости недвижимости с использованием нейронных сетей. Задача оценки недвижимости схематично представлена на рисунке 2.2

Виды нейронных сетей, особенности, отличия
почему выбран именно такой,
сравнение разных способов на основе других статей

Для достижения цели необходимо выбрать факторы, влияющие на рыночную стоимость объектов недвижимости, подготовить выборку для обучения нейронной сети. Обучающая выборка построена для проектирования и обучения нейронной сети с учителем, поскольку такой тип нейронных сетей больше всего подходит для задач, когда имеется большой набор настоящих данных для обучения алгоритма. Исходя из сравнительного анализа нескольких типов нейронных сетей с учителем, проведенного в статье, было решено использовать нейронную сеть многослойный персептрон с использовани-

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|---------------|---------|-----------------|----------|--------------------|---------|--------|-------------|---------------|------------------|---------------|----------|-----------------|---------------|---------------------|-----------------------|------------|
| City | City Id | State | State Id | Building area, sqm | Carport | Garage | Open spaces | Property Type | Property Type Id | Property Type | Bedrooms | Number of sales | Average Price | Median Rental Price | Change in Rental Rate | Sold Price |
| TERRIGAL | 1 | New South Wales | 1245.0 | 1245.0 | 0 | 2 | 0 | 2 | Townhouse | 2 | 3 | 182 | 1087705 | 590 | 7 | 790000 |
| TERRIGAL | 2 | New South Wales | 1102.0 | 1102.0 | 0 | 2 | 0 | 1 | House | 2 | 2 | 53 | 1360472 | 615 | 6 | 751550 |
| TERRIGAL | 3 | New South Wales | 1111.0 | 1111.0 | 0 | 1 | 0 | 2 | Townhouse | 2 | 2 | 182 | 1087705 | 590 | 7 | 642000 |
| DAVISTOWN | 4 | New South Wales | 1330.0 | 1330.0 | 0 | 2 | 1 | 1 | House | 3 | 6 | 38 | 845250 | 490 | -6 | 970000 |
| SARATOGA | 5 | New South Wales | 1260.0 | 1260.0 | 0 | 2 | 0 | 1 | House | 2 | 4 | 87 | 936559 | 490 | 4 | 1007000 |
| KINCUMBER | 6 | New South Wales | 1219.0 | 1219.0 | 0 | 2 | 0 | 1 | House | 2 | 4 | 98 | 741236 | 495 | 15 | 835000 |
| AVOCA BEACH | 7 | New South Wales | 1243.0 | 1243.0 | 0 | 0 | 0 | 1 | House | 2 | 4 | 88 | 1195110 | 650 | 24 | 1220000 |
| KINCUMBER | 8 | New South Wales | 1295.0 | 1295.0 | 0 | 2 | 0 | 1 | House | 3 | 5 | 98 | 741236 | 495 | 15 | 1401000 |
| AVOCA BEACH | 9 | New South Wales | 1368.0 | 1368.0 | 0 | 2 | 0 | 1 | House | 3 | 3 | 88 | 1195110 | 650 | 24 | 1005000 |
| AVOCA BEACH | 10 | New South Wales | 1261.0 | 1261.0 | 0 | 1 | 0 | 1 | House | 2 | 3 | 88 | 1195110 | 650 | 24 | 1015000 |
| AVOCA BEACH | 11 | New South Wales | 1116.0 | 1116.0 | 1 | 0 | 0 | 2 | Townhouse | 2 | 3 | 88 | 1195110 | 650 | 24 | 718000 |
| NORTH GOSFORD | 12 | New South Wales | 196.0 | 196.0 | 0 | 1 | 0 | 4 | Unit | 2 | 2 | 54 | 598092 | 450 | 13 | 520000 |
| AVOCA BEACH | 13 | New South Wales | 1515.0 | 1515.0 | 0 | 3 | 0 | 1 | House | 3 | 5 | 88 | 1195110 | 650 | 24 | 2600000 |
| WYONG | 14 | New South Wales | 1230.0 | 1230.0 | 0 | 2 | 0 | 1 | House | 2 | 3 | 76 | 645917 | 395 | 10 | 710000 |
| BATEAU BAY | 15 | New South Wales | 1318.0 | 1318.0 | 0 | 1 | 0 | 1 | House | 2 | 3 | 197 | 744420 | 485 | 8 | 642500 |
| SARATOGA | 16 | New South Wales | 1313.0 | 1313.0 | 0 | 2 | 0 | 1 | House | 3 | 4 | 87 | 936559 | 490 | 4 | 1512000 |
| YATTALUNGA | 17 | New South Wales | 1220.0 | 1220.0 | 0 | 1 | 0 | 1 | House | 1 | 4 | 9 | 839000 | 410 | 15 | 945000 |
| AVOCA BEACH | 18 | New South Wales | 1205.0 | 1205.0 | 1 | 1 | 0 | 2 | Townhouse | 3 | 3 | 88 | 1195110 | 650 | 24 | 850000 |
| AVOCA BEACH | 19 | New South Wales | 1220.0 | 1220.0 | 0 | 2 | 0 | 1 | House | 3 | 3 | 88 | 1195110 | 650 | 24 | 1400000 |
| SARATOGA | 20 | New South Wales | 1255.0 | 1255.0 | 0 | 2 | 0 | 1 | House | 2 | 3 | 87 | 936559 | 490 | 4 | 880000 |
| GOSFORD | 21 | New South Wales | 1161.0 | 1161.0 | 0 | 2 | 0 | 4 | Unit | 2 | 3 | 5 | 651000 | 368 | -3 | 470000 |
| YATTALUNGA | 22 | New South Wales | 1194.0 | 1194.0 | 0 | 1 | 0 | 1 | House | 2 | 4 | 9 | 839000 | 410 | 15 | 778000 |
| TERRIGAL | 23 | New South Wales | 1135.0 | 1135.0 | 0 | 1 | 0 | 2 | Townhouse | 3 | 3 | 182 | 1087705 | 590 | 7 | 840000 |
| AVOCA BEACH | 24 | New South Wales | 1286.0 | 1286.0 | 0 | 1 | 0 | 1 | House | 2 | 4 | 88 | 1195110 | 650 | 24 | 1510000 |
| TERRIGAL | 25 | New South Wales | 1266.0 | 1266.0 | 0 | 2 | 0 | 1 | House | 3 | 4 | 182 | 1087705 | 590 | 7 | 1025000 |
| AVOCA BEACH | 26 | New South Wales | 1277.0 | 1277.0 | 0 | 2 | 0 | 1 | House | 1 | 2 | 88 | 1195110 | 650 | 24 | 875000 |
| AVOCA BEACH | 27 | New South Wales | 1140.0 | 1140.0 | 0 | 1 | 0 | 4 | Unit | 1 | 2 | 88 | 1195110 | 650 | 24 | 700000 |
| TUMBI UMBI | 28 | New South Wales | 1202.0 | 1202.0 | 0 | 1 | 0 | 1 | House | 2 | 4 | 75 | 813226 | 460 | 3 | 615000 |
| NORTH AVOCA | 29 | New South Wales | 1302.0 | 1302.0 | 0 | 5 | 0 | 1 | House | 3 | 4 | 53 | 1360472 | 615 | 6 | 1000000 |
| KINCUMBER | 30 | New South Wales | 1280.0 | 1280.0 | 0 | 1 | 2 | 1 | House | 3 | 4 | 98 | 741236 | 495 | 15 | 790000 |
| WEST GOSFORD | 31 | New South Wales | 186.0 | 186.0 | 0 | 2 | 0 | 4 | Unit | 1 | 2 | 6 | 605500 | 440 | 10 | 400500 |
| SARATOGA | 32 | New South Wales | 1190.0 | 1190.0 | 0 | 2 | 0 | 1 | House | 2 | 4 | 87 | 936559 | 490 | 4 | 615000 |
| TERRIGAL | 33 | New South Wales | 1265.0 | 1265.0 | 0 | 2 | 0 | 4 | Unit | 2 | 4 | 182 | 1087705 | 590 | 7 | 2120000 |
| EAST GOSFORD | 34 | New South Wales | 1143.0 | 1143.0 | 0 | 1 | 0 | 2 | Townhouse | 2 | 3 | 30 | 742150 | 470 | 9 | 565000 |
| TERRIGAL | 35 | New South Wales | 1475.0 | 1475.0 | 0 | 3 | 0 | 1 | House | 3 | 5 | 182 | 1087705 | 590 | 7 | 1565000 |

Рисунок 2.1 – Исходные данные в csv формате

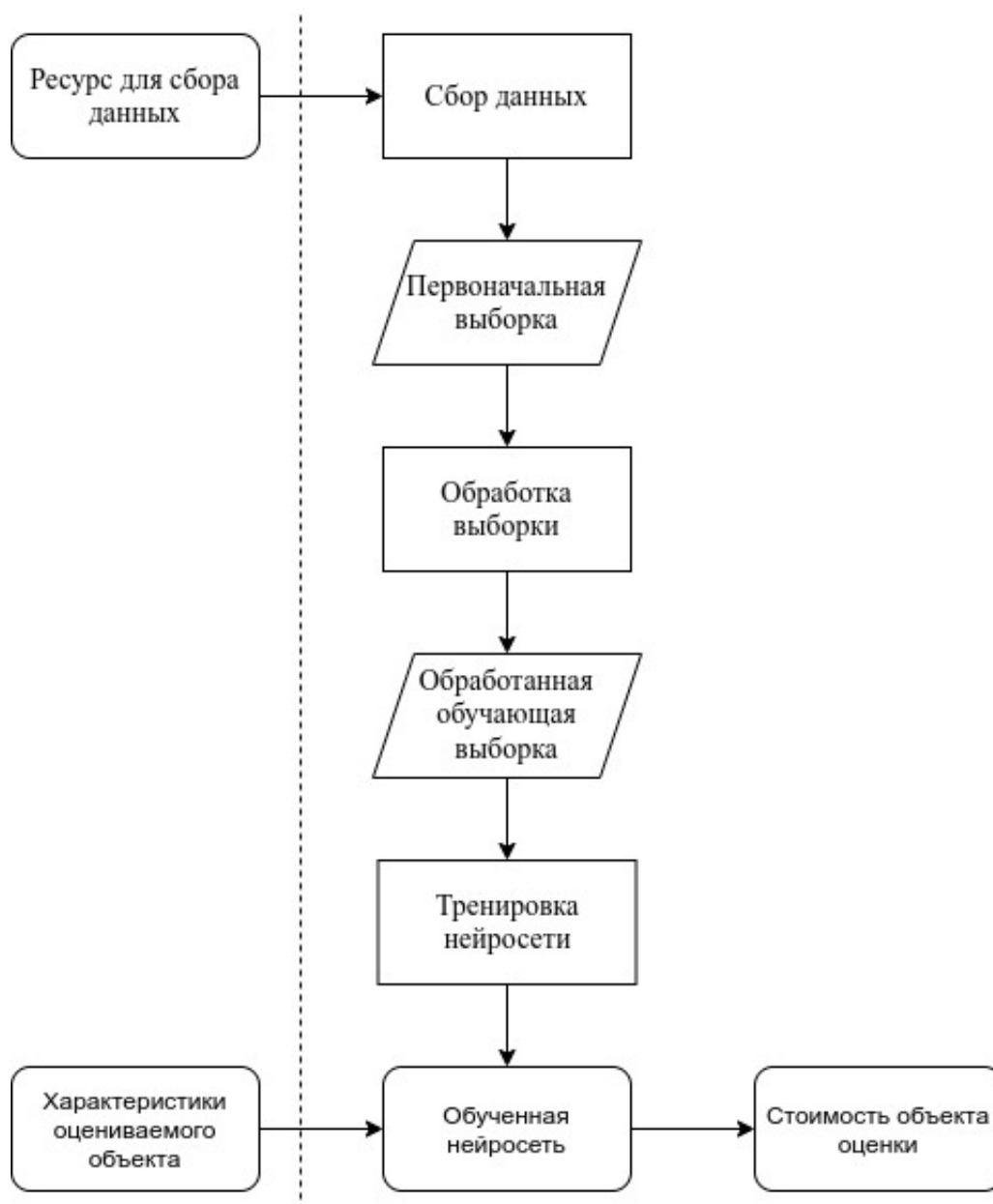


Рисунок 2.2 – Схема использования нейронных сетей для оценки стоимости недвижимости

ем метода обратного распространения ошибки. Многослойным персептроном называют нейронную сеть прямого распространения, где входной сигнал распространяется от слоя к слою в прямом направлении. В общем представлении такая нейронная сеть состоит из:

- множества входных узлов, образующих входной слой;
- одного или нескольких скрытых слоев вычислительных нейронов;
- одного выходного слоя нейронов.

Обобщенная схема многослойного персептрона показана на рисунке 2.3

Алгоритм обратного распространения ошибки является популярным

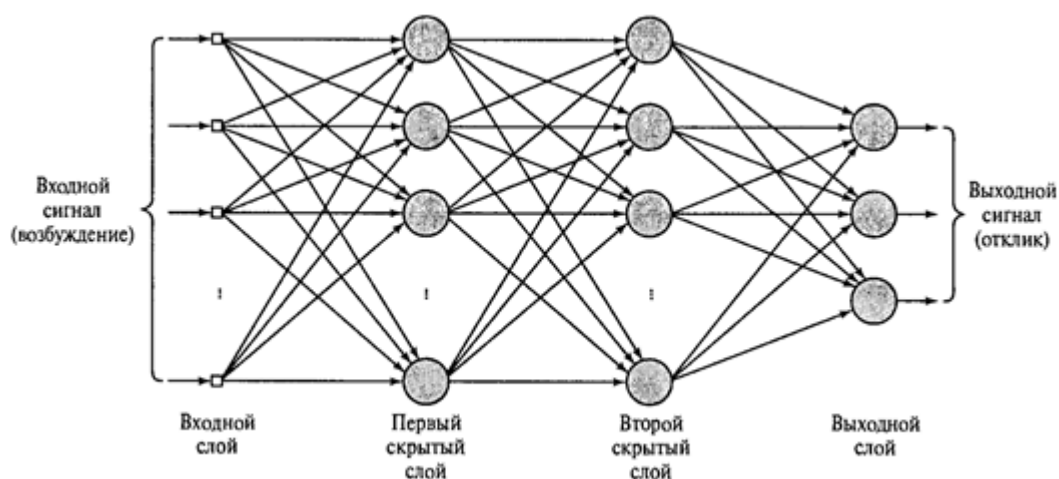


Рисунок 2.3 – Схема многослойного персептрона

алгоритмом обучения нейронных сетей с учителем. В основе идеи алгоритма лежит использование выходной ошибки нейронной сети для вычисления величин коррекции весов нейронов в скрытых слоях

$$E = \frac{1}{2} \sum_{i=1}^k (y - y')^2, \quad (1)$$

где: k — число выходных нейронов сети, y — целевое значение, y' — фактическое выходное значение. Алгоритм является итеративным. На каждой итерации происходит прямой и обратный проходы. На прямом проходе входной вектор распространяется входов сети к ее выходам, в результате формируется выходной вектор, который соответствует фактическому состоянию весов. После вычисляется ошибка нейронной сети как разность между фактическим и целевым значениями. На обратном проходе эта ошибка распространяется от выхода сети к ее входам, и производится коррекция весов нейронов. Полученные данные дают возможность с достаточной точностью прогнозировать стоимость объектов недвижимости по заданным параметрам.

2.4 Сравнение результатов

На основании проведенных исследований можно утверждать, что применение нейронных сетей для прогнозирования стоимости объектов недвижимости более эффективно использования методов регрессионного анализа и может достаточно точно отражать рыночную стоимость недвижимости. Предложенные методы могут быть использованы продавцами для автоматизации оценки стоимости жилой недвижимости, а покупателями могут использоваться в качестве дополнительного источника достоверной и объек-

тивной информации. Кроме того, с использованием предложенных методов моделирования стоимости объектов недвижимости для покупателей становится более прозрачным процесс формирования конечной цены, поскольку ему доступны степени влияния каждой характеристики оцениваемого объекта недвижимости на конечную стоимость.

ЗАКЛЮЧЕНИЕ

В ходе прохождения практики была выполнена программа практики и отражен ход выполнения в дневнике прохождения практики. Все пункты индивидуального задания, данного руководителем практики, были выполнены.

На основании проведенных исследований можно утверждать, что применение нейронных сетей для прогнозирования стоимости объектов недвижимости более эффективно использования методов регрессионного анализа и может достаточно точно отражать рыночную стоимость недвижимости. Предложенные методы могут быть использованы продавцами для автоматизации оценки стоимости жилой недвижимости, а покупателями могут использоваться в качестве дополнительного источника достоверной и объективной информации. Кроме того, с использованием предложенных методов моделирования стоимости объектов недвижимости для покупателей становится более прозрачным процесс формирования конечной цены, поскольку ему доступны степени влияния каждой характеристики оцениваемого объекта недвижимости на конечную стоимость.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

[1] Специфика ценообразования на рынке жилья и факторы, влияющие на цену недвижимости [Электронный ресурс]. — Электронные данные. — Режим доступа: <https://auditfin.com/fin/2009/2/Rodionova/Rodionovadostupa>: 04.10.2020.

[2] Метод наименьших квадратов [Электронный ресурс]. — Электронные данные. — Режим доступа: <https://wiki.loginom.ru/articles/least-squares-method.html>. — Дата доступа: 04.10.2020.

[3] Коэффициент детерминации [Электронный ресурс]. — Электронные данные. — Режим доступа: <https://wiki.loginom.ru/articles/coefficient-of-determination.html>. — Дата доступа: 04.10.2020.

[4] Аппроксимация [Электронный ресурс]. — Электронные данные. — Режим доступа: <https://wiki.loginom.ru/articles/approximation.html>. — Дата доступа: 04.10.2020.

[5] Регрессионный анализ [Электронный ресурс]. — Электронные данные. — Режим доступа: <https://wiki.loginom.ru/articles/regression-analysis.html>. — Дата доступа: 04.10.2020.