

Анкета участника

Информация	Описание
ФИО студента	Медведева Татьяна Васильевна
Направление/специальность	080800.62 «Прикладная информатика (бакалавриат)»
Вуз	Государственное образовательное учреждение высшего профессионального образования Нижегородский государственный архитектурно-строительный университет
Вуз-партнер	Да
Город	Нижний Новгород
Кафедра	Информационные системы в экономике
ФИО зав. кафедрой	Папкина М.Д.
Тема ВКР	Применение современных информационных технологий и интеллектуальных методов анализа в задаче оценки недвижимости
Версия Deductor	5.1
Дата защиты	23.06.2010
Оценка	Отлично
Руководитель ВКР	Канд. ф.-м. наук, доцент Прокопенко Наталья Юрьевна
Представлено	<ul style="list-style-type: none"> ▪ Анкета участника ▪ Аннотация ▪ Пояснительная записка ▪ Сканированные титульные листы ▪ Рецензия ▪ Отзыв научного руководителя ▪ Справка об использовании ▪ Сценарий Deductor (с данными) ▪ Презентация Power Point

Аннотация

ПРИМЕНЕНИЕ СОВРЕМЕННЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И ИНТЕЛЛЕКТУАЛЬНЫХ МЕТОДОВ АНАЛИЗА В ЗАДАЧЕ ОЦЕНКИ НЕДВИЖИМОСТИ

Среди элементов рыночной экономики недвижимость занимает особое место, выступая как в качестве средств производства, так и в качестве объекта потребления. Поэтому определение рыночной стоимости объектов недвижимого имущества является одним из основных этапов при принятии управленческих решений государственными органами управления и риэлтерскими компаниями.

В данном дипломном проекте описываются современные подходы к решению задач оценки недвижимости.

Объектом исследования данной работы являются теоретические и практические проблемы процесса оценки недвижимости с использованием методов статистического и интеллектуального анализа данных.

Предметом исследования выбраны методы и результаты оценки объектов вторичного рынка жилья г. Нижнего Новгорода.

Информационной базой для решения задачи оценки недвижимости являются данные риэлтерских компаний о продаже однокомнатных квартир вторичного рынка недвижимости Приокского района г. Нижнего Новгорода, материалы периодической печати по изучаемой проблеме и web-сайты недвижимости. Данные были собраны с веб-сайта агентства недвижимости Волгожилстрой, с федерального портала о недвижимости «Мир квартир», с печатного и электронного вариантов газеты «Из рук в руки».

Дипломный проект содержит введение, две главы, заключение, список использованной литературы, три приложения, графический и табличный материалы.

В первой главе дипломного проекта излагаются теоретические аспекты оценки рынка недвижимости, а также дается анализ состояния рынка недвижимости города Нижнего Новгорода за последние 10 лет.

На основе указанных особенностей рынка вторичного жилья, представленной динамики стоимости вторичного жилья за период январь 2000-июнь 2010 г.г., изученной структуры распределения квартир по критериям (районность, комнатность, тип планировки, материал стен), делается вывод, что состояние рынка недвижимости г. Нижнего Новгорода сильно изменилось на протяжении последних десяти лет.

В первой главе также описаны существующие три подхода к определению рыночной стоимости недвижимости: 1) подход с точки зрения затрат; 2) оценка по прямому сравнению продаж; 3) подход с точки зрения доходности, исследованы факторы, влияющие на стоимость недвижимости. Отмечено, что для решения задачи оценки недвижимости, перечисленные подходы реализуются с помощью различных экспертных методов, а также методов математической статистики, методов Data Mining.

С точки зрения оценки недвижимости актуальными являются две основные задачи – классификация объектов недвижимости и оценка стоимости жилой недвижимости. В первой главе дается обзор методов прогнозирования и классификации для их дальнейшего практического использования при решении задач оценки недвижимости. Подробно описываются наиболее популярные методы, основанные на модели множественной регрессии, модели дерева решений и модели нейронной сети.

В условиях невозможности получения информации необходимого качества или ее высокой зашумленности, неполноте и противоречивости эффективно работают модели, основанные на современных инновационных методах, в том числе на методах интеллектуального анализа данных таких, как нейронные сети и деревья решений.

Интеллектуальный анализ данных невозможно реализовать без специализированных пакетов программ, реализующих алгоритмы выявления закономерностей, решения задач классификации и прогнозирования. В качестве инструментального средства поддержки принятия решений в области оценки недвижимости на основе интеллектуальных средств была выбрана аналитическая платформа Deductor Academic. Главными критериями выбора являются:

- наличие бесплатной учебной версии Deductor Academic;
- наличие самых современных методов извлечения манипулирования, визуализации данных, кластеризации, прогнозирования и многих других технологий интеллектуального анализа данных в Deductor Academic.

Во второй главе дипломного проекта рассмотрено практическое применение современных информационных технологий и интеллектуальных методов анализа на примере задачи оценки стоимости жилья, проведен анализ статистических данных работы процесса для его оптимизации и повышения конкурентоспособности работы агентства недвижимости.

При помощи аналитической платформы Deductor построен сценарий, решающий поставленные в первой главе задачи классификации и прогнозирования стоимости жилья. Задачи классификации решены при помощи моделей дерева решений и нейронной сети, задачи прогнозирования – при помощи модели нейронной сети и модели множественной регрессии. Также было проведено сравнение модели на основе показателей оценки их качества.

Первым шагом создания сценария является загрузка данных с помощью мастера импорта. Учитывая ограниченные возможности работы с информацией версии Deductor Academic, данные, хранятся в файле Kvartiriy.txt. При загрузке информации необходимо правильно установить соответствия между входными данными и значениями по умолчанию.

В данной работе для выявления дубликатов, противоречий, аномальных значений в загруженных данных о продажах квартир были использованы следующие обработчики:

- Дубликаты и противоречия;
- Диаграмма;
- Фильтрация данных.

С помощью обработчика данных «Дубликаты и противоречия» в исходной выборке данных выявлены дублирующие и противоречивые записи.

Следующим шагом этапа очистки и предобработки было изучение аномальных значений. Выявление аномальных значений проводилось с помощью визуализатора «Диаграмма». Для этого вначале с помощью обработчика «Калькулятор» была вычислена цена за 1 кв.м. и построен график этого показателя, предварительно ранжированного по возрастанию.

Визуальный анализ позволил выявить крайние точки, где график себя ведет неоднородно.

Экцессы были исключены с помощью обработчика «Фильтрация» (было выявлено 24 аномальных значения).

После применения процедур очистки был сделан вывод, что данные готовы к построению моделей.

Цель дальнейшего исследования состояла в построении модели классификации, относящей исследуемую квартиру к одному из трех классов «дорогие, средние, дешевые квартиры». Для решения задачи классификации использовался метод деревьев решений. Алгоритм построения деревьев решений позволяет решать задачи классификации, и при этом легко интерпретировать результаты, выдаваемые построенной моделью. Было построено дерево решений, на вход которого подавались характеристики квартиры, а выходом служили 3 класса: дорогие, средние, дешевые квартиры.

В аналитической платформе Deductor имеется обработчик «Дерево решения», который работает только с дискретными значениями, поэтому было принято решение разбить все данные на три категории, в зависимости от цены за 1 кв.м. Это было сделано с помощью обработчика данных «Квантование».

Квантование – это процесс, в результате которого происходит распределение значений непрерывных данных по конечному числу интервалов заданной длины.

Для задания параметров квантования в списке полей было выделено поле Цена за 1 кв. м., которое и подвергалось разбиению. В результате весь диапазон был разбит на 3 интервала. Следующий шаг - подключение обработчика «Замена значений». В результате выполнения данного обработчика производится замена значений по таблице подстановок, которая содержит пары, состоящие из исходного значения и выходного значения. Например, <Цена за 1 кв.м. до 60000 > – <Дешевая квартира>, <Цена за 1 кв.м. от 60000 до 62857,15 > – <Средняя квартира>, <Цена за 1 кв.м. от 62857,15 > – <Дорогая квартира>. Данное разбиение и замена значений являются предварительными этапами для построения дерева решения.

Результатом работы Обработчика «Дерево решений» является список иерархических правил вида «Если.., то», пример приведен на Рис. 1.

На основании полученных правил определяем класс, к которому относиться любая квартира: «Дешевая квартира», «Средняя квартира», «Дорогая квартира». Например, для того, чтобы определить тип квартиры, можно рассмотреть правило 3: если одновременно: Остальная площадь меньше 20,5 м², Жилая площадь больше или равно 14 м² и Кухня больше или равно 5,4 м², то квартира, имеющая такие параметры, будет отнесена к классу средних квартир (рис.1).

Правил: 5 из 5		Фильтр: Без фильтрации		Следс		Поддержка		Достоверность		
№	Номер правила	Показатель	Знак	Значение	ab	Цена з	Кол-во	%	Кол-во	%
1	1	9.0 Остальная площадь	<	20,5	дешевая	з	2	2,50	2	100,00
		9.0 Жилая площадь	<	14						
2	2	9.0 Остальная площадь	<	20,5	дорогая	з	21	26,25	17	80,95
		9.0 Жилая площадь	>=	14						
		9.0 Кухня	<	5,4						
3	3	9.0 Остальная площадь	<	20,5	средняя	з	50	62,50	26	52,00
		9.0 Жилая площадь	>=	14						
		9.0 Кухня	>=	5,4						
		9.0 Жилая площадь	<	20,5						
4	4	9.0 Остальная площадь	<	20,5	дешевая	з	2	2,50	2	100,00
		9.0 Жилая площадь	>=	14						
		9.0 Кухня	>=	5,4						
5	5	9.0 Жилая площадь	>=	20,5	дешевая	з	5	6,25	5	100,00
		9.0 Остальная площадь	>=	20,5						

Рис. 1. Пример набора правил

Поддержка показывает, какой процент сделок из имеющегося набора данных удовлетворяет условию правила, а достоверность показывает, какой процент сделок, удовлетворяющих условию правила, удовлетворяет и его следствию (с какой вероятностью оцениваемая квартира, удовлетворяющая условию правила, удовлетворяет и его следствию).

В сценарии было построено несколько моделей классификации на основе Деревя решений с разным порядком уровня доверия.. Используя визуализатор Диаграмма рассеяния, было выявлено, что большая точность классификации достигается при уровне доверия 20%. С целью получения более компактного дерева, для более простой интерпретации полученных правил, уровень доверия был снижен до 5% (при этом качество этой модели ухудшилось).

Задача определения класса квартир объектов недвижимости решалась также при помощи модели нейронной сети.

В аналитической платформе Deductor существует специальный обработчик «Нейросеть». Для того чтобы построить модель классификации на основе нейронной сети, необходимо выполнить предварительную обработку данных, а именно применить обработчики «Квантование» и

«Замена значений» (аналогично тому как это было сделано в классификаторе «Дерево решений»).

Для решения задачи классификации была выбрана следующая архитектура нейронной сети: входной слой состоял из 9 нейронов, на которые подавались значения девяти известных факторов (Жилая площадь, Кухня, Стены, Этаж, Балкон, Санузел, Телефон, Состояние, Остальная площадь), один скрытый слой состоял из двух нейронов и выходной слой состоял из трех нейронов, соответствующих трем классам (Дешевая, Средняя, Дорогая квартира).

После обучения нейронной сети при помощи визуализатора «Что-если» имеется возможность проверить, как работает построенный нейросетевой классификатор. А именно с помощью данного визуализатора покупатель или риелтор, может определить класс, к которому относиться квартира с заданными параметрами.

Прежде чем строить модель прогнозирования на основе множественной регрессии, был проведен корреляционный анализ для выявления и устранения мультиколлинеарности факторов, оценки зависимости выходного поля от входных факторов и устранения незначимых факторов (в Deductor имеется обработчик «Корреляционный анализ»). Принцип корреляционного анализа состоит в поиске таких значений, которые в наименьшей степени взаимосвязаны с выходным результатом. Такие факторы могут быть исключены из результирующего набора данных практически без потери полезной информации. Критерием принятия решения об исключении является порог значимости. Если корреляция (степень взаимозависимости) между входным и выходным факторами меньше порога значимости, то соответствующий фактор отбрасывается как незначимый. После проведения корреляционного анализа становится доступным визуализатор «Матрица корреляции». В работе представлена матрица корреляции (а именно, результат объединения нескольких таблиц с помощью обработчика Слияние) для исходных данных. После устранения незначимых факторов, была повторно построена матрица корреляции и сделан вывод, что оставшиеся факторы попарно неколлинеарны.

Для выявления значимости атрибутов, которые в большей степени определяют стоимость квартиры был использован визуализатор «Значимость атрибутов». В результате было установлено, что наибольшее влияние на цену оказывают три атрибута: Жилая Площадь – 29%, Кухня – 28% и Остальная Площадь – 22%.

Для определения стоимости квартиры на основании отобранных факторов, используя возможности Deductor, была построена модель множественной линейной регрессии. Коэффициенты линейной регрессии были получены при помощи обработчика данных «Линейная регрессия». Тестирование модели было проведено на новых данных. Сравнение спрогнозированной цены и фактической было проведено с помощью построения графика.

Для решения задачи оценки стоимости жилья также была построена нейросетевая модель.

Работа нейронной сети аналогична работе эксперта, который может оценить стоимость объекта недвижимости только на основе его свойств (признаков).

Первым шагом в построении нейросетевой модели, позволяющей рассчитывать стоимость недвижимости, был выбор архитектуры нейросети: входной и скрытый слой такие же, как в нейросетевом классификаторе, выходной слой имеет один нейрон, где будем получать прогнозируемое значение цены. Обучение нейронной сети проводилось при помощи обработчика «Нейросеть». Прогнозируемую цену клиент или риелтор могут определить при помощи визуализатора «Что-если». Например, Если Клиент задает желаемые параметры: то на выходе получает прогнозируемую стоимость квартиры. (рис. 2).

Поле	Значение
Входные	
9.0 Жилая площадь	18
9.0 Кухня	5
9.0 Стены	4
9.0 Этаж	1
9.0 Балкон	1
9.0 С/У	1
9.0 Телефон	1
9.0 Состояние	4
9.0 Остальная площ...	25
Выходные	
9.0 Цена	2624049,58604281

Рис. 2. Пример применение визуализатора «Что-если» для определения цены

Для качественного применения методов прогнозирования и классификации необходимо построить несколько моделей, оценить эффективность каждой из них и выбрать лучшую, то есть требуется провести сравнение моделей.

Для определения качества модели «Дерево решений» и модели «Нейронная сеть» в задаче классификации недвижимости использовался специальный визуализатор в аналитической платформе Deductor «Таблица сопряженности», который позволяет наглядно оценить результаты классификации. Так как процент правильно распознанных примеров модели, построенной на основе нейронной сети (72%) выше процента правильно распознанных примеров модели, построенной на основе дерева решений (64%) был сделан вывод, что при решении задачи классификации целесообразно использовать модель нейронной сети, так как она дает меньшую ошибку.

Для решения задачи прогнозирования были построены модель множественной регрессии и модель нейронной сети. Оценка качества моделей свидетельствует о том, что полученные модели сравнимы по точности, так как имеют процент правильно распознанных примеров 79% и 74% соответственно.

Вследствие того, что нейронные сети представляют собой обучающиеся модели, что позволяет «доучивать» их при поступлении новой информации, в работе делается вывод, что при решении задачи прогнозирования на новых данных целесообразно использовать модель нейронной сети.