

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.89

Петрусевич
Максим Сергеевич

**Программное средство поиска и аренды жилья с использованием
технологии Apache Cayenne**

АВТОРЕФЕРАТ

на соискание степени

магистра информатики и вычислительной техники
по специальности 1-40 81 01 – **Информатика и технологии разработки
программного обеспечения**

Научный руководитель
Смолякова Ольга Георгиевна
к.т.н.

Минск 2020

ВВЕДЕНИЕ

Развитие программных средств тесно связано с повседневной жизнью современного человека. Всё больше и больше рутинных действий люди выполняют при помощи компьютеров и автоматизированных систем. Не обошла автоматизация и сферу торговли. С древних времён в сфере торговли требовалось совершать множество рутинных действий для донесения покупателю информации о товаре, организации продажи, сбора информации о продаваемых товарах и многое другое. Развитие технической науки позволило сократить время на проведение рутинных операций и упростить хранение информации. Развитие технологий позволило максимально упростить и ускорить все рабочие моменты, которые ранее занимали большое количество времени.

Одной из областей торговли является сфера аренды и продажи недвижимости. Данная сфера также, как и другие сферы торговли, проходила непростые этапы ручной обработки информации, когда каждую новую поступающую информацию требовалось обрабатывать людям, и на такую обработку уходило множество времени. Однако неавтоматизированность сферы приносила неудобства всем субъектам сделки. Раньше, для того чтобы любому человеку разместить объявление о продаже недвижимости, необходимо было отправить письмо в редакцию местной газеты, что занимало значительное время.

Автоматизация в сфере торговли, также пришедшая и в сферу продажи недвижимости, позволила сократить затрачиваемое время как редакций, выкладывающих объявления, так и пользователей, которые эти объявления подавали. Теперь, посредством использования компьютеров, вычислительных систем и сетей, пользователям не обязательно совершать долгие по времени действия, достаточно лишь наличие персонального компьютера и связи с сетью интернет. Редакции также смогли упростить свою работу и переложить хранение, обработку и выдачу информации на различные приложения, которые взаимодействуют с пользователями.

Однако, несмотря на активное развитие рынка автоматизированной продажи недвижимости, достаточно сложно найти универсальную и модульную систему, которая смогла бы полностью удовлетворять запросам пользователей, т.к. с ростом автоматизации многие пользователи стали более требовательны к используемым ими ресурсам. Как правило, существующие отечественные аналоги, реализующие функционал интернет-портала по продаже, покупке и аренде недвижимости, не полностью реализуют весь потенциал автоматиза-

ции продаж. Например, в современных отечественных системах отсутствует возможность пользователей просматривать статистические данные о районах на картах, как это сделано, например, в зарубежных аналогах. Таким образом, возникает естественная ниша в среде программных средств, реализующих функционал биржи по продаже, покупке и аренде недвижимости, на создание сервиса, который мог бы ещё больше упростить и ускорить поиск необходимой недвижимости пользователями.

Целью данной диссертации является разработка клиент-серверного программного средства, автоматизирующего некоторые процессы работы биржи недвижимости с возможностью просмотра статистической информации.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Современные технологии, применяемые в сфере недвижимости, позволяют производить анализ рынка продажи и аренды недвижимости используя множество методов, включая методы обработки больших объёмов статистических данных. Однако не всегда покупатель или профессиональный риелтор могут в полной мере воспользоваться собранными данными, т.к. порой закономерности влияние тех или иных факторов не являются очевидными.

В связи с вышеописанным, актуальной темой является разработка и исследование возможности использования в сфере продажи и аренды недвижимости нейронных сетей с собранными и обработанными статистическими данными.

Степень разработанности проблемы

Аналогами разрабатываемого программного средства являются крупнейшие агрегаторы в сфере недвижимости в Беларуси (kufar.by) и зарубежом (trulia.com). Однако стоит отметить, что на данных ресурсах не используются метода машинного обучения, как инструменты анализа статистической информации. Можно судить о недостаточной разработанности в связи с отсутствием данного метода применения нейронных сетей в сфере недвижимости.

Разработанное в ходе написания магистерской диссертации программное средство и подходы направлены на устранение одного из главных недостатков существующих аналогов - невозможности анализа предоставленной информации.

Цель и задачи исследования

Целью данной магистерской диссертации является разработка программного средства (далее - ПС), предназначенного для упрощения поиска жилья, его аренды, продажи или покупки с использованием нейронных сетей и технологий обработки данных.

В результате анализа предметной области и рынка аналогов разрабатываемого программного средства были определены основные цели и функционал, которые должны быть реализованы. Функциональные и нефункциональные требования, определённые в результате анализа предметной области, были описаны в диаграммах и разделах магистерской диссертации.

Основываясь на анализе аналогов разрабатываемого программного средства, была поставлена задача создать ПС с отличными от аналогов уникаль-

ными возможностями (функциями, характеристиками), которые позволили бы однозначно идентифицировать программное средство на рынке сервисов, предоставляющих доступ к поиску, покупке, продаже или аренде жилья в Республике Беларусь.

Архитектура приложения представлена в виде отдельных модулей, каждый из которых решает определённые задачи сбора, обработки, выдачи и отображения данных, используемых конечным пользователем. Используемые технологии были выбраны исходя из наиболее подходящих для решения определённых задач.

Объектом исследования магистерской диссертации является использование нейронных сетей и технологий обработки данных в сфере недвижимости.

Предметом исследования является процесс обучения нейронной сети для предсказания цен на недвижимость на основе статистической информации о районах города Минска.

Основной гипотезой, положенной в основу диссертационной работы, является возможность использования нейронных сетей в совокупности со статистической информацией для предсказания реальной цены на недвижимость. Среди такой информации была выбрана статистика о близлежащих школах, детских садах, транспортном сообщении и средней цене по городу на уже продаваемую недвижимость. В настоящее время аналоги программных средств, работающих со статистикой по недвижимости, не предоставляют возможности использования нейронных сетей в этой области. Если созданная модель нейронной сети сможет показывать лучший результат, чем профессиональные риелторы - это может избавить рынок недвижимости от переоценённых квартир и упростить их поиск для конечных пользователей.

Теоретическая и методологическая основа исследования

В основу диссертации легли работы исследователей, публицистов и учёных из области машинного обучения и создания программного обеспечения. Информационная база сформирована на основе литературы, открытой информации, технических нормативно-правовых актов, а также сведений из электронных ресурсов.

Научная новизна

Научная новизна и значимость полученных результатов работы заключается в разработке программного средства с использованием машинного обучения и статистической информации.

Теоретическая значимость заключается в анализе воздействия статистических факторов на ценообразование на рынке недвижимости.

Практической значимостью полученных результатов можно считать упрощение поиска жилья конечными пользователями, экономическая польза - возможность убрать лишнее звено продавец-покупатель в сфере недвижимости, избавившись таким образом от дополнительных трат.

Основные положения, выносимые на защиту

1. Автоматизация сбора, агрегации и анализа статистической информации позволит сформировать целостную структуру данных, которые в будущем можно будет использовать для применения их в нейронных сетях.

2. Данные подаваемые на вход нейронной сети могут сформировать настроенные параметры, которые в будущем могут быть использованы для предсказания цены на недвижимость.

3. Сформированная нейронная сеть сможет делать предсказания по цене проданной недвижимости более точно, чем профессиональный риелтор.

Апробация диссертации и информация об использовании ее результатов

Отдельные положения диссертации могут быть использованы при преподавании дисциплин связанных с машинным обучением и анализом статистических данных.

Публикации

По теме диссертации опубликована 1 печатная работы в рецензируемом издании. Общий объем публикаций по теме диссертации составляет 12 страниц.

Структура и объем работы

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения, списка использованных источников, списка публикация автора и приложений. В первой главе представлен анализ технологии разработки, прототипов и формирование требований к проектируемому программному средству. Во второй главе представлен анализ требований к программному средству и разработка функциональных требований. В третьей - проектирование и разработка программного средства. В четвёртой главе представлено использование нейронной сети в предметной области. В приложениях к диссертации представлены исходные коды программного средства и скрипты баз данных.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** рассматриваются предпосылки использования автоматизирующих средств в сфере недвижимости. В данной главе освещено как и какие инструменты были использованы в разные исторические этапы существования сферы недвижимости.

В **первой главе** производится анализ существующих аналогов разрабатываемого программного средства, рассматриваются их положительные и отрицательные стороны. Далее, на основе полученных сведений, производится анализ литературных источников по данным темам, к таким темам относятся, например, обработка больших объёмов информации, машинное обучение и литература, связанная с проектированием программных средств.

Во **второй главе** производится анализ требований к программному средству и разработка функциональных требований. Более точная постановка требований выполняется за счёт анализа процесса поиска недвижимости.

Для того, чтобы определить, какие именно этапы и шаги нуждаются в автоматизации, требуется составить первоначальный алгоритм поиска недвижимости, который используют клиенты при поиске недвижимости. Поиск жилья состоит из следующих этапов:

- а) определение площадки поиска, будь то печатные издания или интернет-ресурсы;
- б) поиск среди имеющихся на площадке выставленных объявлений, подходящих по критерию цены;
- в) отсеивание среди найденных объявлений неподходящих по определённым вторичным критериям: метраж жилой площади, количество комнат, этаж, год постройки дома;
- г) личная оценка найденного района, в том числе оценка криминальной составляющей района;
- д) оценка транспортного сообщения района;
- е) поиск и оценка аналогичных продаваемых площадей во избежание переплаты;
- ж) связь с продавцом;
- з) осмотр недвижимости;
- и) принятие решения.

Описанные выше этапы представляют собой алгоритм поиска недвижимости с полным анализом многих критериев, которые могут повлиять, в конечном итоге, на удовлетворённость покупкой и уровень жизни клиента. Как

видно из описанного алгоритма, многие этапы представляют собой во многом анализ тех или иных критериев, которые ложатся на плечи клиента.

Одним из важных этапов в алгоритме нахождения недвижимости является личная оценка найденного района. Данный этап представляет собой оценку районов, в которых находится искомая недвижимость, на основе личных предпочтений клиента. Данный подход к анализу является субъективным, т.к. пользователь не имеет каких-либо точных и достоверных данных. Однако, клиент может использовать сторонние ресурсы для проведения анализа и оценки района. Это позволит более реально оценить состояние интересующего района по фактору криминальности и наличия хорошего транспортного сообщения, однако, во многих случаях, ручной поиск всей необходимой информации может занять достаточно долгое время. Таким образом, имеет смысл в агрегировании всей информации в разрабатываемом программном средстве для сокращения временных затрат пользователей на анализ. Поиск и оценка аналогичных продаваемых площадей во избежание переплаты; отсеивание среди найденных объявлений неподходящих по определённым вторичным критериям: метраж жилой площади, количество комнат, этаж, год постройки дома и оценка транспортного сообщения района имеют схожий алгоритм действий. Как и в описанной ранее личной оценке найденного района, данные этапы предполагают использование двух подходов: субъективной оценки клиента на основе личного опыта или формализованного подхода с поиском информации на сторонних ресурсах.

Описанные выше этапы представляют собой формализованный поиск дополнительной информации, которая может быть полезна пользователю в последующем выборе недвижимости. Таким образом, существует потребность в создании единого сервиса, который мог бы упростить и ускорить поиск недвижимости со стороны клиента.

В третьей главе описаны этапы проектирования и разработки программного средства с учётом поставленных требований и построенных функциональных моделей.

В четвёртой главе диссертации описаны методы машинного обучения, которые были использованы при настройке гиперпараметров модели нейронной сети, а также обзор влияния статистических данных на работу нейронной сети. В данной главе происходит исследование и анализ влияния заранее обработанной статистики.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Были собраны статистические данные связанные со сферой недвижимости в городе Минске. Сбор данной статистики показал, что в Республике Беларусь на данный момент отсутствуют централизованные источники информации, которые могут быть использованы в сферах автоматизации.

2. Создано программное средство с использованием систем обработки данных и инструментов машинного обучения. Во время разработки был выработан подход к получению, обработке и работе со статистическими данными в полной схеме работы программного средства, использующего нейронные сети.

3. Было определено влияние статистических параметров на ценообразование в сфере недвижимости.

Рекомендации по практическому использованию результатов

Полученные результаты могут быть внедрены в системы используемые риелторами при продаже недвижимости, а также в системы-агрегаторы, для отображения спрогнозированной цены пользователям системы.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

Статьи в рецензируемых изданиях:

1-А. Петрусевич, М.С. Статистическое исследование сервиса StackOverflow / М.В. Всяких – Белгород: Международный научный журнал «Научные вести», 2019, 12(17). – с. 357–365.