

Исследование прогностической модели рынка цифровых фотоаппаратов

Чипизубов Кирилл Михайлович

Приамурский государственный университет имени Шолом-Алейхема

Студент

Баженов Руслан Иванович

Приамурский государственный университет им. Шолом-Алейхема

к.п.н., доцент, зав.кафедрой информационных систем, математики и методик обучения

Аннотация

В данной статье рассматриваются методы регрессионного анализа в среде «GretL» для разработки и построения прогностической модели, а также прогнозирование с помощью нейронной сети реализованной в программе «Neural Network Wizard». Исследование производится на основе данных рынка поддержанных цифровых фотоаппаратов. Экспериментальные данные были взяты с сайта www.avito.ru. В результате исследования были получены модели, позволяющие с очень низкой точностью спрогнозировать стоимость покупки, либо продажи поддержанного цифрового фотоаппарата в зависимости от заданных критериев. Низкая точность прогноза объясняется тем, что область исследования обладает огромным ценовым диапазоном, несовершенной разработанной прогностической моделью и маленьким объёмом исходных данных, не позволяющих добиться желаемого результата.

Ключевые слова: регрессионный анализ, нейронная сеть, рынок цифровых фотоаппаратов;

Research of the forecast model of the digital cameras market

Chipizubov Kirill Mikhailovich

Sholom-Aleichem Priamursky State University

Student

Bazhenov Ruslan Ivanovich

Sholom-Aleichem Priamursky State University

Candidate of pedagogical sciences, associate professor, Head of the Department of Information Systems, Mathematics and teaching methods

Abstract

In this article, one shows the regression analysis methods in the “GretL” environment meant for development and building a prognostic model and also making a prediction by means of neural network realized by «Neural Network Wizard» software. The research uses data taken from the used digital cameras

market. The experimental data were taken from www.avito.ru. As a result one received models which can be poorly used to predict the buy/sale price of a used digital camera depending of given criteria. The prediction's low quality can be explained by the fact that the field of study contains a huge price range, incomplete prognostic model and small amount of input data, which does not allow to reach a desired result.

Keywords: regression analysis, neural network, the market for digital cameras

Бурно развивающаяся отрасль разработки мобильных телефонов, позволяет иметь всегда под рукой возможность оперативно сделать необходимый снимок. Но устройства данного типа далеко не во всех случаях имеют возможность удовлетворить потребности в качестве сделанных фотографий, т.к. они не специализированы на данной функции, а лишь имеют её в целях расширения функционала и дополнительного критерия для конкурентирования и завлечения покупателей, уделяющих ей особое внимание.

Для создания фотографий высокого качества используют цифровые профессиональные фотоаппараты, имеющие специализированные параметры, модули и характеристики направленные на максимально приближённые к идеальному решению поставленной задачи.

В связи с тем, что покупка нового цифрового фотоаппарата с магазина не является дешёвой и легкодоступной для многих людей, планирующих использовать данное устройство в любительских целях, были взяты в качестве основы для моделирования и исследования, данные с сайта *avito.ru* по Еврейской автономной области и городу Хабаровск. В качестве ключевых параметров были взяты: число эффективных пикселей, размер дисплея фотоаппарата в дюймах и цена устройства.

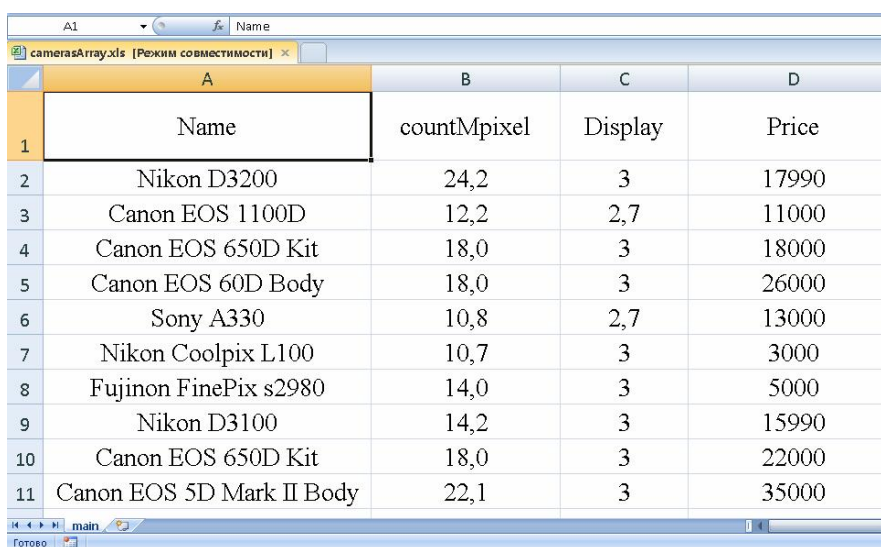
Нейросетевые технологии в банковской деятельности изучала Т.А. Серебрякова [1]. Нейронные сети и их применения рассматривали А.П.Филиппов, А.В.Щагин [2]. Разработкой интеллектуальной карты занимался Р.И.Баженов [3]. Информационная поддержка синтеза схмотехнически интегрированных принципиальных электрических схем изучал В.Г. Манжула [4]. Нейросетевые технологии на финансовом рынке рассматривал С. Д. Кулик и А. Д. Жижилев [5]. Место модульных нейронных сетей в классификации искусственных нейронных сетей показывал Н.С. Костин [6]. В.А.Седов и Н.А.Седова применили моделирование расхождения судов в зоне чрезмерного сближения нейронными сетями [7]. Использование априорной информации о месторасположении объектов лица для оценки его биометрических признаков нейронной сетью исследовал А.Д.Варламов [8].

Первым шагом к прогнозированию цены цифровых фотоаппаратов является сбор и структуризация данных предметной области для разработки и исследования регрессионной модели с помощью прикладного программного пакета GretL и табличного процессора Excel из пакета Microsoft Office 2007.

В связи с тем, что покупка нового цифрового фотоаппарата с магазина не является дешёвой и легкодоступной для многих людей, планирующих

использовать данное устройство в любительских целях, были взяты в качестве основы для моделирования и исследования, данные с сайта *avito.ru* по Еврейской автономной области и городу Хабаровск. В качестве ключевых параметров были взяты: число эффективных пикселей, размер дисплея фотоаппарата в дюймах и цена устройства.

Для удобства ввода, редактирования и структурированного отображения собираемых данных, было решено составлять таблицу в табличном процессоре Excel пакета Microsoft Office 2007. На завершении этапа сбора данных, файл с данными имел следующий вид (рис. 1). А также был сохранён в формат совместимости с Excel 97-2003.



| | A | B | C | D |
|----|---------------------------|-------------|---------|-------|
| 1 | Name | countMpixel | Display | Price |
| 2 | Nikon D3200 | 24,2 | 3 | 17990 |
| 3 | Canon EOS 1100D | 12,2 | 2,7 | 11000 |
| 4 | Canon EOS 650D Kit | 18,0 | 3 | 18000 |
| 5 | Canon EOS 60D Body | 18,0 | 3 | 26000 |
| 6 | Sony A330 | 10,8 | 2,7 | 13000 |
| 7 | Nikon Coolpix L100 | 10,7 | 3 | 3000 |
| 8 | Fujinon FinePix s2980 | 14,0 | 3 | 5000 |
| 9 | Nikon D3100 | 14,2 | 3 | 15990 |
| 10 | Canon EOS 650D Kit | 18,0 | 3 | 22000 |
| 11 | Canon EOS 5D Mark II Body | 22,1 | 3 | 35000 |

Рисунок 1. Исходные данные, представленные в MS Excel 2007

Имея таблицу с исходными данными, начинаем разработку регрессионной модели для исследования с помощью прикладного программного пакета GretL, импортировав в него наш .xls-файл (рис. 2), нажав сочетание клавиш «CTRL+O» для вызова диалогового окна выбора импортируемого файла. Также в фильтре отображаемых типов файлов выбрать необходимый нам тип.

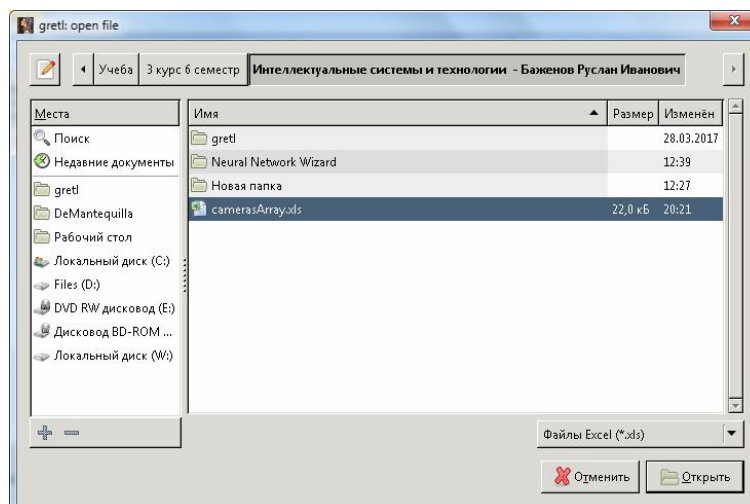


Рисунок 2. OpenDialog прикладного программного продукта "GretL"

После выбора файла с данными, открывается основная форма “GretL” со списком названий полей идентичных названиям столбцов исходного файла (рис. 3).

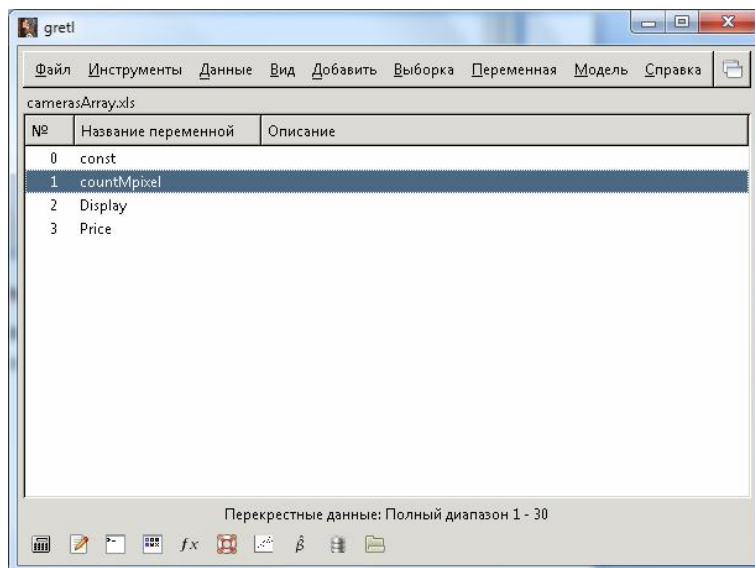


Рисунок 3. MainForm GretL с импортированными данными

Далее формируем нашу регрессионную модель, используя метод наименьших квадратов. Для этого выбираем пункт главного меню «Модель →Метод наименьших квадратов» (рис. 4).

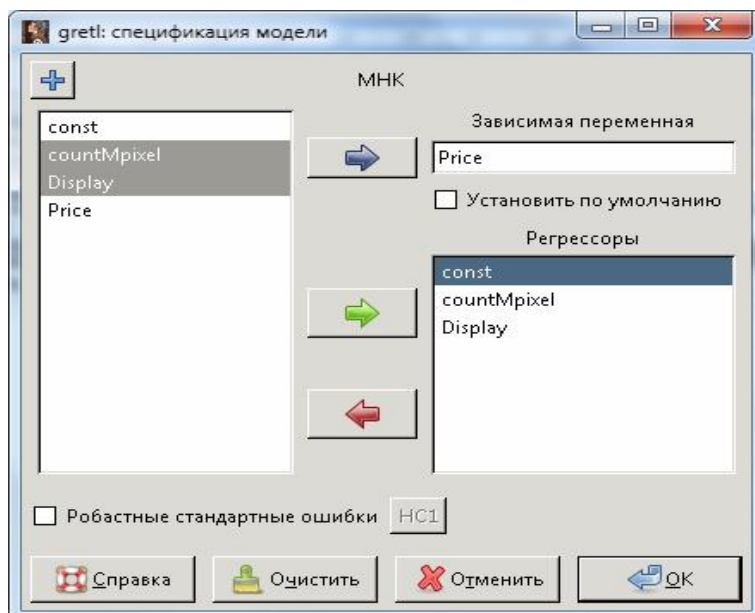


Рисунок 4. Окно спецификации модели

В данном окне (рис. 4) выставляем зависимые и входные переменные, влияющие на создаваемую модель и её характеристики. После нажатия на кнопку «ОК» откроется форма отображения сформированной регрессионной модели (рис. 5).

| | Коэффициент | Ст. ошибка | t-статистика | P-значение |
|-------------|-------------|------------|--------------|------------|
| const | -77902,7 | 41388,1 | -1,882 | 0,0706 * |
| countMpixel | 1489,08 | 602,618 | 2,471 | 0,0201 ** |
| Display | 24617,0 | 16070,7 | 1,532 | 0,1372 |

| | | | |
|----------------------|-----------|------------------------|----------|
| Среднее зав. перемен | 21322,67 | Ст. откл. зав. перемен | 14648,09 |
| Сумма кв. остатков | 3,34e+09 | Ст. ошибка модели | 11120,99 |
| R-квадрат | 0,463351 | Испр. R-квадрат | 0,423599 |
| F(2, 27) | 11,65609 | P-значение (F) | 0,000224 |
| Лог. правдоподобие | -320,4854 | Крит. Акаике | 646,9709 |
| Крит. Шварца | 651,1745 | Крит. Хенна-Куинна | 648,3156 |

Рисунок 5. Форма отображения сформированной модели

Значение показателя «Испр. R-квадрат» указывает на вероятность успешного прогноза модели. Выходит, что шанс на точный прогноз равен всего 42%. Это не очень хороший показатель для прогнозирования.

Чтобы попробовать исправить столь неприятную ситуацию попробуем прологарифмировать переменную Price. Для этого закроем окно сформированной модели, выделим, нажатием интересующую нас переменную и выберем пункт главного меню «Добавить → Логарифмы выделенных переменных» (рис. 6).

| № | Название переменной | Описание |
|---|---------------------|----------------------|
| 0 | const | |
| 1 | countMpixel | |
| 2 | Display | |
| 3 | Price | |
| 4 | l_Price | = логарифм для Price |

Рисунок 6. Добавление логарифмированной переменной

Теперь заново формируем регрессионную модель рынка фотоаппаратов с помощью метода наименьших квадратов, но в качестве зависимой переменной указываем l_Price (рис. 7).

После нажатия на кнопку «ОК», на открывшейся форме сформированной модели (рис. 8), смотрим и сравниваем значения «Испр. R-квадрат» текущей и предыдущей модели. В обоих случаях целая часть значения показателя равна 42%. Числа различаются на 0,3% в пользу первой модели.

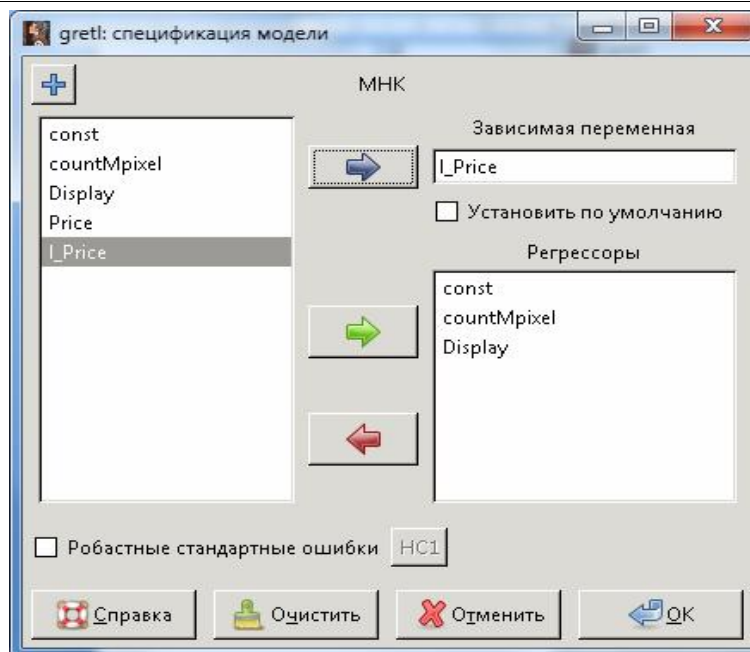


Рисунок 7. Окно спецификации модели с прологарифмированной зависимой переменной

| | Коэффициент | Ст. ошибка | t-статистика | P-значение |
|-------------|-------------|------------|--------------|------------|
| const | 6,30520 | 2,22778 | 2,830 | 0,0087 *** |
| countMpixel | 0,104083 | 0,0324368 | 3,209 | 0,0034 *** |
| Display | 0,527571 | 0,865027 | 0,6099 | 0,5470 |

| | | | |
|----------------------|-----------|------------------------|----------|
| Среднее зав. перемен | 9,713216 | Ст. откл. зав. перемен | 0,786143 |
| Сумма кв. остатков | 9,674785 | Ст. ошибка модели | 0,598603 |
| R-квадрат | 0,460191 | Испр. R-квадрат | 0,420206 |
| F(2, 27) | 11,50887 | P-значение (F) | 0,000243 |
| Лог. правдоподобие | -25,59304 | Крит. Акаике | 57,18608 |
| Крит. Шварца | 61,38967 | Крит. Хеннана-Куинна | 58,53085 |

Рисунок 8. Окно сформированной модели с прологарифмированной зависимой переменной

Результаты исследования регрессионной модели рынка цифровых фотоаппаратов, с помощью прикладного программного пакета “GretL” показали нам, что прогнозировать данную область рынка с таким набором данных крайне не удачная идея, но цель данной статьи – «Разработка прогностической модели рынка цифровых фотоаппаратов», поэтому создадим ещё одну книгу MS Excel 2007 и назовём её «Регрессионная модель.xls». После этого откроем её и введём коэффициенты, полученные путём исследования исходных данных в GreatL. Добавляем столбец «Требуемые параметры» и результирующую ячейку «Цена». В итоге была

создана прогностическая модель разработанная на основе исследований регрессионной модели (Рис. 9).

| | A | B | C | D | E |
|---|-------------|---------|---|---------------------|-----|
| 1 | Const | 6,3052 | | Требуемые параметры | |
| 2 | countMpixel | 0,10408 | | Мегапиксели | 21 |
| 3 | Display | 0,52757 | | Размер дисплея | 2,8 |
| 4 | | | | | |
| 5 | l_Price | 9,96814 | | | |
| 6 | | | | | |
| 7 | Цена | 21335,8 | | | |

Рисунок 9. Прогностическая модель, разработанная в MS Excel

Следующим этапом нашего исследования является разработка прогностической модели рынка цифровых фотоаппаратов с помощью нейронных сетей. В качестве эмулятора нейросети будем использовать программу Neural Network Wizard.

Для того, чтобы перенести исходные данные из формата «xls» в формат «txt», с которым работает Neural Network Wizard, открываем наш файл с исходными данными и сохраняем его в формате «txt» с помощью команды «Меню→Сохранить как...» (Рис. 10).

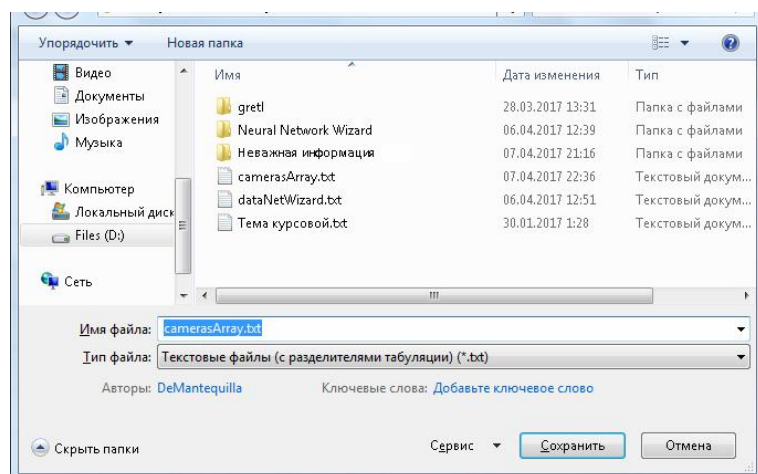


Рисунок 10. Сохранение исходных данных в текстовом формате

После сохранения файла, необходимо открыть его в текстовом редакторе и создать структуру с помощью табуляции (рис. 11).

Откроем «Neural Network Wizard» и, подгрузив отформатированный текстовый файл, нажмём кнопку «Далее».

В следующем окне необходимо назначить роль переменных. Для переменной «Name» установим значение «Не использовать» в CheckBox «Использовать поле как». Переменные «countMpixel» и «Display» назначим входными, а «Price» целевой (рис. 12).

| | Name | countMpixel | Display | Price |
|----|---------------------------|-------------|---------|-------|
| 1 | | | | |
| 2 | Nikon D3200 | 24.2 | 3 | 17990 |
| 3 | Canon EOS 1100D | 12.2 | 2.7 | 11000 |
| 4 | Canon EOS 650D Kit | 18.0 | 3 | 18000 |
| 5 | Canon EOS 60D Body | 18.0 | 3 | 26000 |
| 6 | Sony A330 | 10.8 | 2.7 | 13000 |
| 7 | Nikon Coolpix L100 | 10.7 | 3 | 3000 |
| 8 | Fujinon FinePix s2980 | 14.0 | 3 | 5000 |
| 9 | Nikon D3100 | 14.2 | 3 | 15990 |
| 10 | Canon EOS 650D Kit | 18.0 | 3 | 22000 |
| 11 | Canon EOS 5D Mark II Body | 22.1 | 3 | 35000 |
| 12 | Sony DSC-HX300 | 20.4 | 3 | 15000 |
| 13 | Nikon D600 Body | 24.3 | 3.2 | 60000 |
| 14 | Canon EOS 600D Kit | 18.0 | 3 | 16000 |
| 15 | Nikon D7100 Kit | 24.1 | 3.2 | 50000 |
| 16 | Canon PowerShot SX420 IS | 20.0 | 3 | 15500 |
| 17 | Canon EOS 450D | 12.2 | 3 | 10000 |
| 18 | Sony Alpha SLT-A58 Kit | 20.1 | 2.7 | 15000 |
| 19 | Canon EOS 550D Kit | 18.0 | 3 | 15000 |
| 20 | Canon EOS 60D Body | 18.0 | 3 | 25000 |
| 21 | Canon EOS 1000D kit | 10.1 | 2.5 | 4000 |
| 22 | Fujifilm X-T10 Body | 16.3 | 3 | 33000 |
| 23 | Canon EOS 60D Body | 18.0 | 3 | 35000 |
| 24 | Canon EOS 7D Kit | 18.0 | 3 | 12000 |
| 25 | Nikon D7100 Kit | 24.1 | 3.2 | 53000 |
| 26 | Canon EOS 600D Kit | 18.0 | 3 | 20500 |
| 27 | Nikon D80 Body | 10.2 | 2.5 | 9000 |
| 28 | Nikon coolpix L810 | 16.1 | 3 | 2700 |
| 29 | Nikon D3200 | 24.2 | 3 | 17000 |
| 30 | Nikon D5200 Kit | 24.1 | 3 | 27000 |
| 31 | Fujifilm X-E2 Body | 16.3 | 3 | 38000 |
| 32 | | | | |

Рисунок 11. Структурированные данные, открытые в программе Notepad++

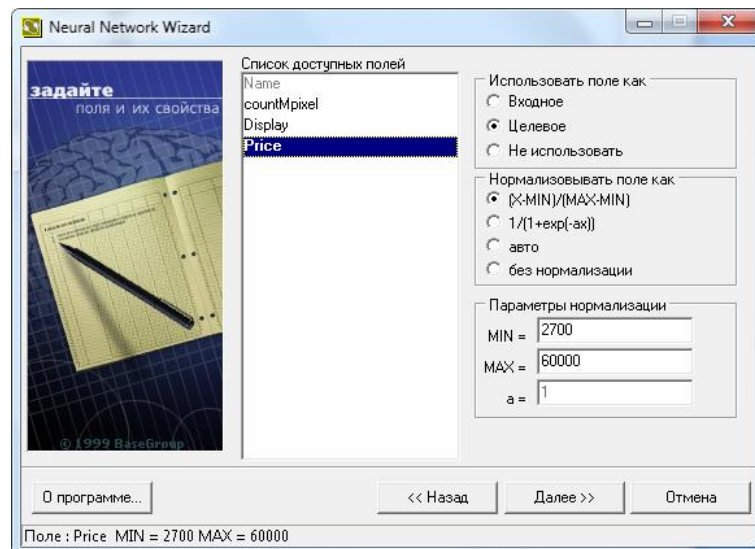


Рисунок 12. Назначение ролей полям модели в Neural Network Wizard

Нажимая кнопку «Далее», попадаем на форму указания параметров прогнозирования модели (рис.13). Оставляем параметры по умолчанию.

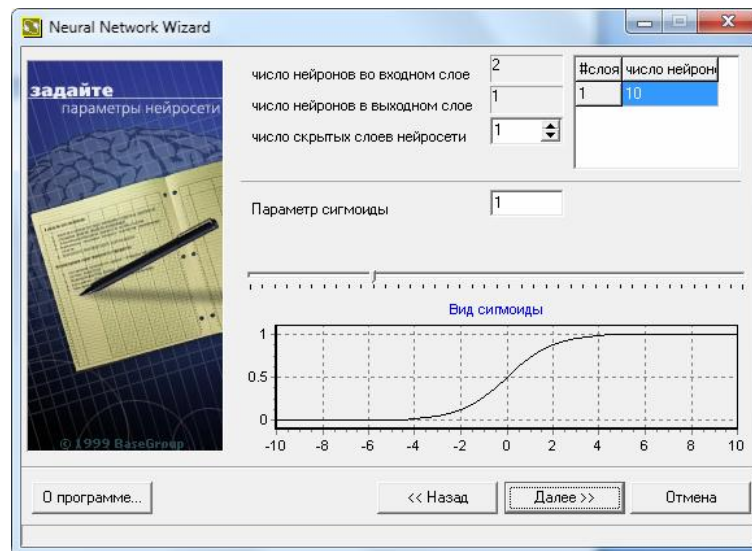


Рисунок 13. Параметры нейросети

Нажимаем кнопку «Далее». На следующей форме необходимо указать параметры обучения нейронной сети (рис. 14). Установим галочку «Прошло 10000 шагов». Также установим значение «100%» для поля «Использовать для обучения нейросети».

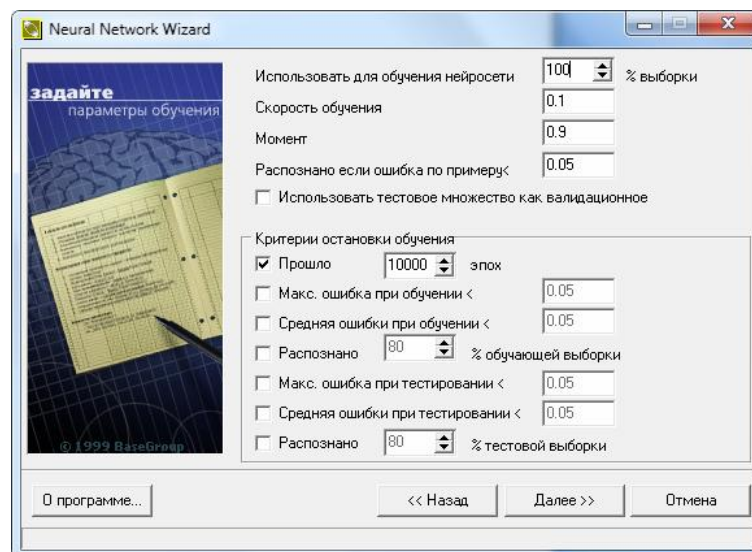


Рисунок 14. Окно параметров обучения нейронной сети

Нажав на кнопку «Далее» мы попадаем на форму со всеми параметрами нейронной сети, которую мы настраивали для построения прогностической модели рынка цифровых фотоаппаратов. После просмотра полного обзора установленных параметров исследования нейронной сетью, нажимаем кнопку «Далее».

В окне обучения нейронной сети (рис. 15) нажимаем кнопку «Пуск обучения» и ожидаем завершения процесса обучения до заданного ранее условия.

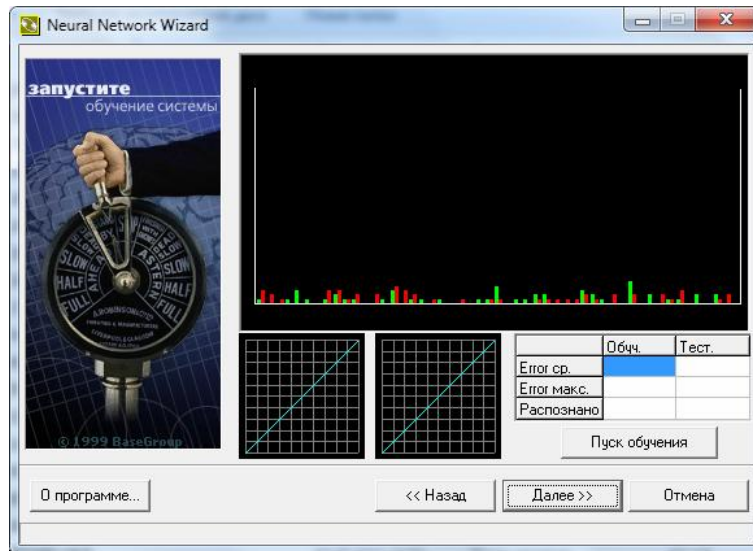


Рисунок 15. Окно обучения нейронной сети

После окончания обучения программа выдаст сообщение о завершении процесса обучения (рис 16).

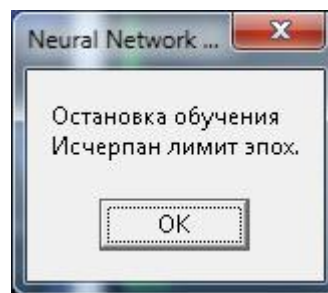


Рисунок 16. Сообщение о завершении обучения

Затем снова нажимаем кнопку «Далее».

Перед нами появляется окно прогнозирования, смоделированной прогностической модели (рис. 17).

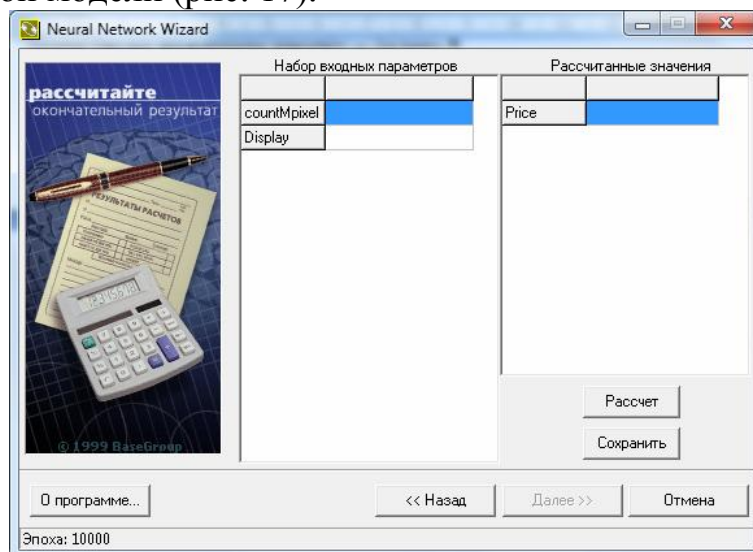


Рисунок 17. Окно прогнозирования

Для проверки прогноза попробуем ввести входные параметры, максимально приближённые к исходным, но не существующие в исходной таблице.

Переменной «countMpixel» присвоим значение “17,9”, а переменной «Display» - “2,9” и нажмём кнопку «Расчёт».

Сравниваем полученные данные в поле «Расчитанные значения» с имеющимися данными в таблице (рис 18).

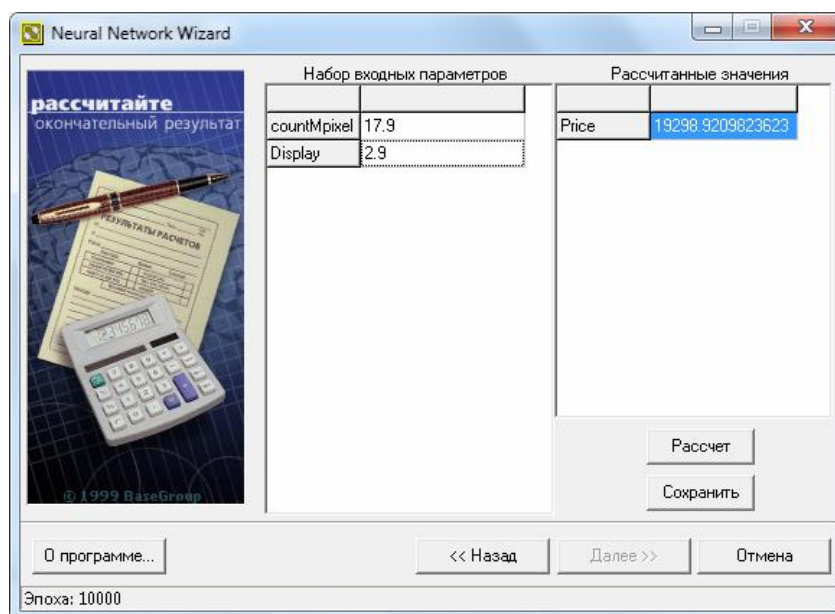


Рисунок 18. Сравнение данных

Далее представлена табл.1 для сравнений результатов прогнозирования прогностической модели разработанной на основе регрессионной модели и прогнозирования с помощью нейронных сетей.

Таблица 1. Сравнение результатов

| Регрессионный анализ | Нейронная сеть | Абсолютная разница | Относительная разница, % |
|----------------------|----------------|--------------------|--------------------------|
| 20058,48 | 19298,92 | 759,56 | 3,79 |
| 15765,59 | 17327,74 | 1562,14 | 9,91 |
| 9958,57 | 21115,11 | 11156,54 | 112,03 |
| 33427,35 | 21549,94 | 11877,41 | 35,53 |
| 21335,80 | 15049,56 | 6286,23 | 29,46 |

В ходе исследования регрессионной модели рынка цифровых фотоаппаратов бывшего употребления оказалось, что успешное

прогнозирование в данной сфере рынка маловероятно, что объясняется огромным ценовым диапазоном, несовершенной разработанной прогностической моделью и недостаточным объёмом исходных данных.

Однако если развивать данный проект, поиском идеальных параметров прогностической модели, дополнением данными исходной таблицы, и постоянным обучением нейросети, то далеко не исключено, что данная модель сможет с очень высокой вероятностью прогнозировать цену на покупку/продажу цифровых фотоаппаратов на территории Еврейской автономной области и города Хабаровск. Также необходимо добавить дополнительные входные данные, которые помогут увеличить качество прогноза спроектированной моделью.

В случае доведения данной модели до состояния максимально приближённого к идеальному, можно разработать программное изделие для помощи людям, которые планируют покупку/продажу цифрового фотоаппарата. А если приложение будет достаточно популярным, то можно расширить область исследования до исследований ценового диапазона цифровых фотоаппаратов по всей стране и стран СНГ.

Библиографический список

1. Серебрякова Т.А. Нейросетивые технологии в банковской деятельности//Актуальные вопросы экономических наук. 2013. №35. С. 183-186.
2. Филиппов А.П., Шагин А.В. Нейронные сети и их применения//национальное исследования университета «МИЭТ». 2016. №23. С.85-88.
3. Круглик Р.И., Баженов Р.И. Разработка интеллектуальной карты // Постулат. 2016. № 12 (14). С.47
4. Манжула В.Г. Информационная поддержка синтеза схемотехнически интегрированных принципиальных электрических схем // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление. 2012. Т. 2. № 145. С. 144-150.
5. Кулик С. Д, Жижилев А. Д. Нейросетевые технологии на финансовом рынке // Нейрокомпьютеры: разработка, применение. 2009. №4. С. 59-78.
6. Костин Н.С. Место модульных нейронных сетей в классификации искусственных нейронных сетей // Интеллектуальный потенциал XXI века: ступени познания. 2013. №19. С. 91-95.
7. Седов В.А., Седова Н.А. Моделирование расхождения судов в зоне чрезмерного сближения нейронными сетями // Научные проблемы транспорта Сибири и Дальнего Востока. 2014. №3. С. 102-105.
8. Варламов А.Д. Использование априорной информации о месторасположении объектов лица для оценки его биометрических признаков нейронной сетью // Нейрокомпьютеры: разработка, применение. 2015. № 3. С. 34-38