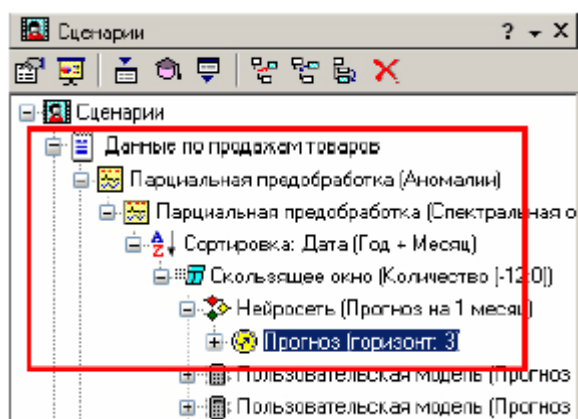


## ЛАБОРАТОРНАЯ РАБОТА 9 ПРОГНОЗИРОВАНИЕ С ПОМОЩЬЮ НЕЙРОННОЙ СЕТИ

Прогнозирование результата на определенное время вперед, основываясь на данных за прошедшее время, – задача, встречающаяся довольно часто. К примеру, перед большинством торговых фирм стоит задача оптимизации складских запасов, для решения которой требуется знать, что и сколько должно быть продано через неделю и т.п., задача предсказания стоимости акций какого-нибудь предприятия через день и т.д. и другие подобные вопросы. **Deductor Studio** предлагает для этого инструмент "Прогнозирование".

Прогнозирование появляется в списке Мастера обработки только после построения какой-либо модели прогноза: нейросети, линейной регрессии и т.д. Прогнозировать на несколько шагов вперед имеет смысл только временной ряд (к примеру, если есть данные по недельным суммам продаж за определенный период, можно спрогнозировать сумму продаж на две недели вперед). Поскольку при построении модели прогноза необходимо учитывать много факторов (зависимость результата от данных день, два, три, четыре назад), то методика имеет свои особенности. Покажем ее на примере.

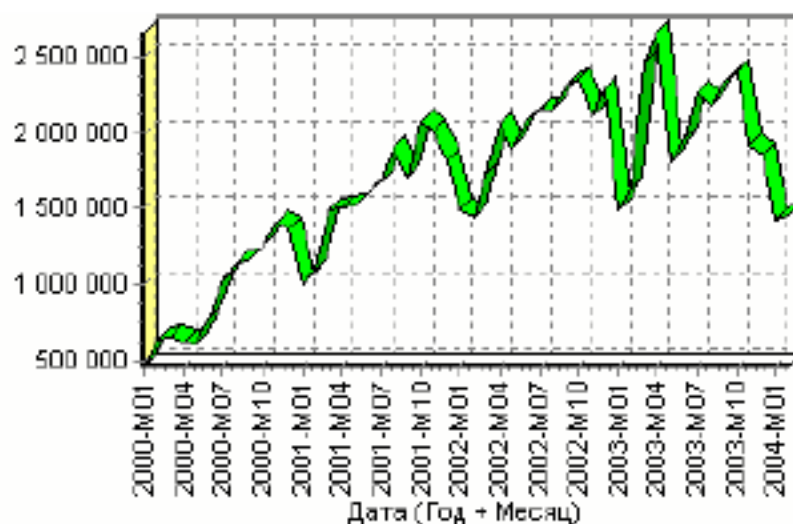
Данный пример анализа можно посмотреть в проекте "Демопример анализа данных.ded".



### Исходные данные

У аналитика имеются данные о ежемесячном количестве проданного товара за несколько лет. Ему необходимо, основываясь на этих данных, определить, какое Количество товара будет продано через месяц и через два.

Исходные данные по продажам находятся в файле "Trade.txt", известному по предыдущему примеру (расчет автокорреляции). Выполним импорт данных из файла, не забыв указать в Мастере, чтобы в качестве разделителя дробной и целой частей была точка, а не запятая.



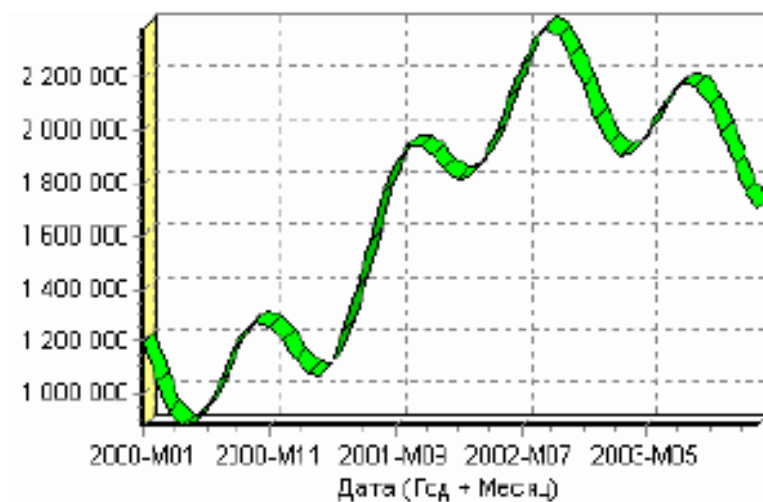
### Удаление аномалий и сглаживание

После импорта данных воспользуемся диаграммой для их просмотра.

На ней видно, что данные содержат аномалии (выбросы) и шумы, за которыми трудно разглядеть тенденцию. Поэтому перед прогнозированием необходимо удалить аномалии и сгладить данные.

Сделать это можно при помощи парциальной обработки. Запустим Мастер обработки, выберем в качестве обработки данных парциальную обработку и перейдем на следующий шаг Мастера. Как известно, второй шаг Мастера отвечает за обработку пропущенных значений, которых в исходных данных нет. Поэтому здесь ничего не настраиваем. Следующий шаг отвечает за удаление аномалий из исходного набора. Выберем поле для обработки "Количество" и укажем для него обработку аномальных явлений (степень подавления – малая).

Четвертый шаг Мастера позволяет провести спектральную обработку. Из исходных данных необходимо исключить шумы, поэтому выбираем столбец "Количество" и указываем способ обработки "Вычитание шума" (степень вычитания – малая). На следующем шаге запустим обработку, нажав на "Пуск". После обработки посмотрим полученный результат на диаграмме.



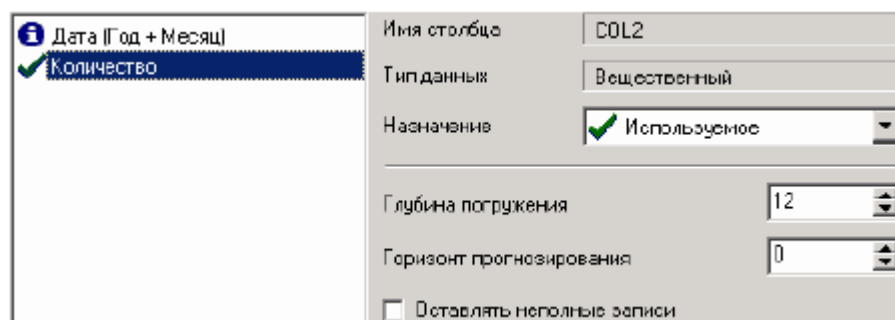
Видно, что данные сгладились, аномалии и шумы исчезли. Также видна тенденция.

Теперь перед аналитиком встает вопрос, а как, собственно, прогнозировать временной ряд. Во всех предыдущих примерах мы сталкивались с ситуацией, когда есть входные столбцы - факторы и есть выходные столбцы - результат. В данном случае столбец один. Строить прогноз на будущее будем, основываясь на данных прошлых периодов, т. е. предполагая, что количество продаж на следующий месяц зависит от количества продаж за предыдущие месяцы. Это значит, что входными факторами для модели могут быть продажи за текущий месяц, продажи за месяц ранее и т.д., а результатом должны быть продажи за следующий месяц, т. е. здесь явно необходимо трансформировать данные к скользящему окну.

### **Скользящее окно 12 месяцев назад**

Запустим Мастер обработки, выберем в качестве обработчика скользящее окно и перейдем на следующий шаг.

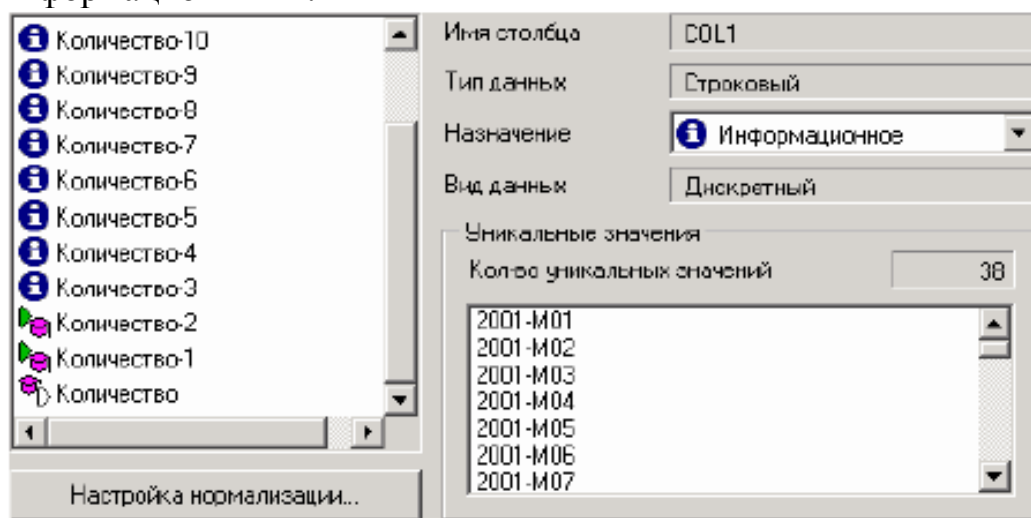
Проведя авторегрессионный анализ, обнаружили наличие годовой сезонности. В связи с этим было решено строить прогноз на месяц вперед, основываясь на данных за 1, 2, 11 и 12 месяцев назад. Поэтому требуется выбрать глубину погружения 12, назначив поле "Количество" используемым. Тогда данные трансформируются к скользящему окну так, что аналитику будут доступны все нужные факторы для построения прогноза.



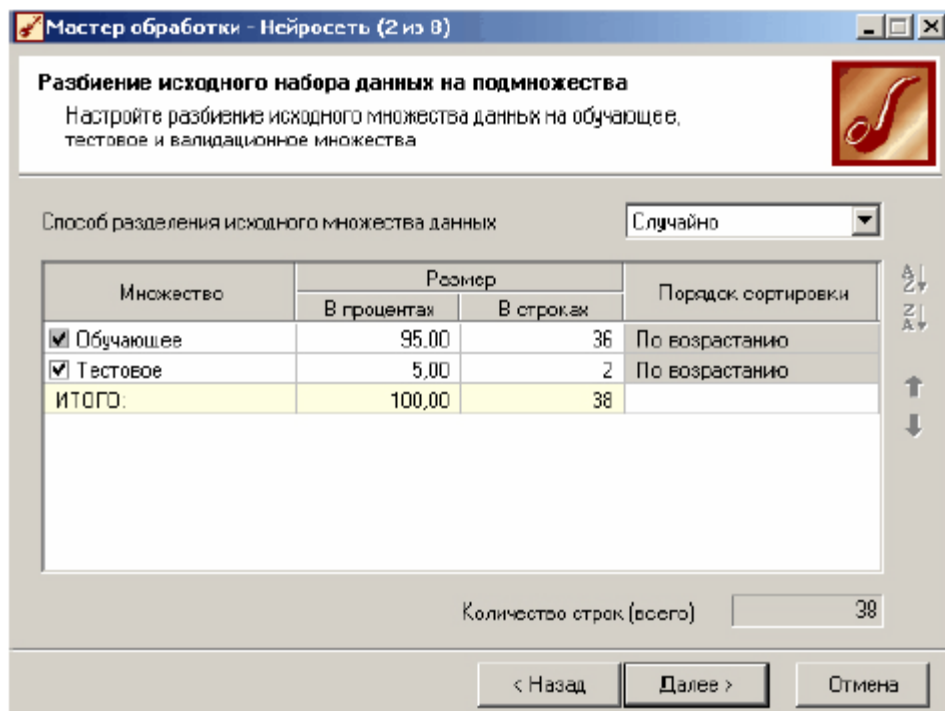
Теперь в качестве входных факторов можно использовать "Количество - 12", "Количество - 11" - данные по количеству 12 и 11 месяцев назад (относительно прогнозируемого месяца), а также "Количество - 2" и "Количество - 1" - данные за 2 предыдущих месяца. В качестве выходного поля укажем столбец "Количество".

### Обучение нейросети (прогноз на 1 месяц вперед)

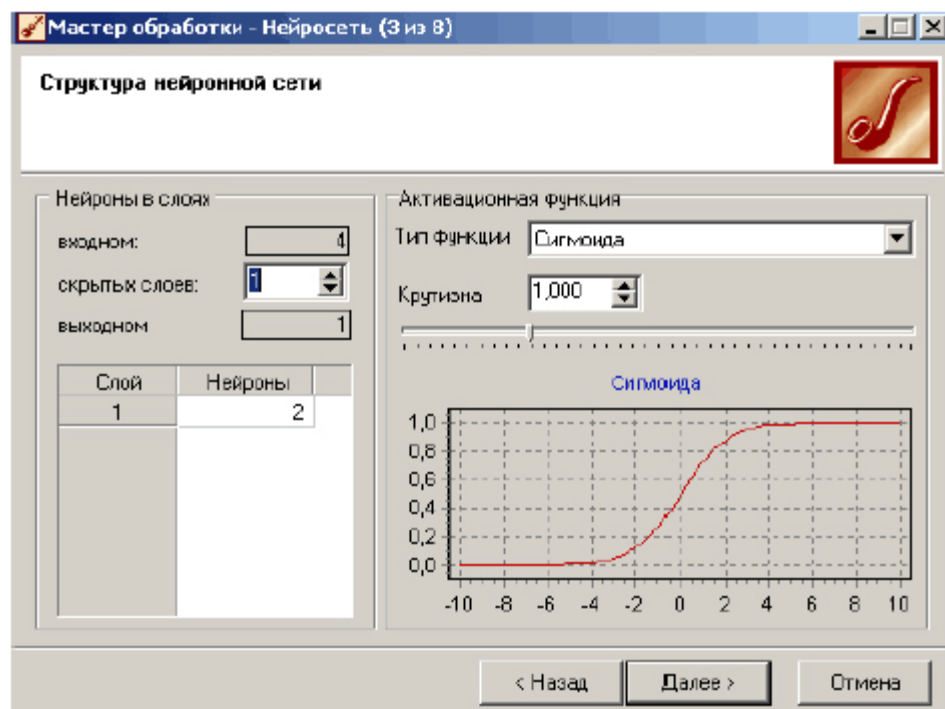
Перейдем непосредственно к самому построению модели прогноза. Откроем Мастер обработки и выберем в нем нейронную сеть. На втором шаге Мастера согласно с принятым ранее решением установим в качестве входных поля "Количество - 12", "Количество - 11", "Количество - 2" и "Количество - 1", а в качестве выходного - "Количество". Остальные поля сделаем информационными.



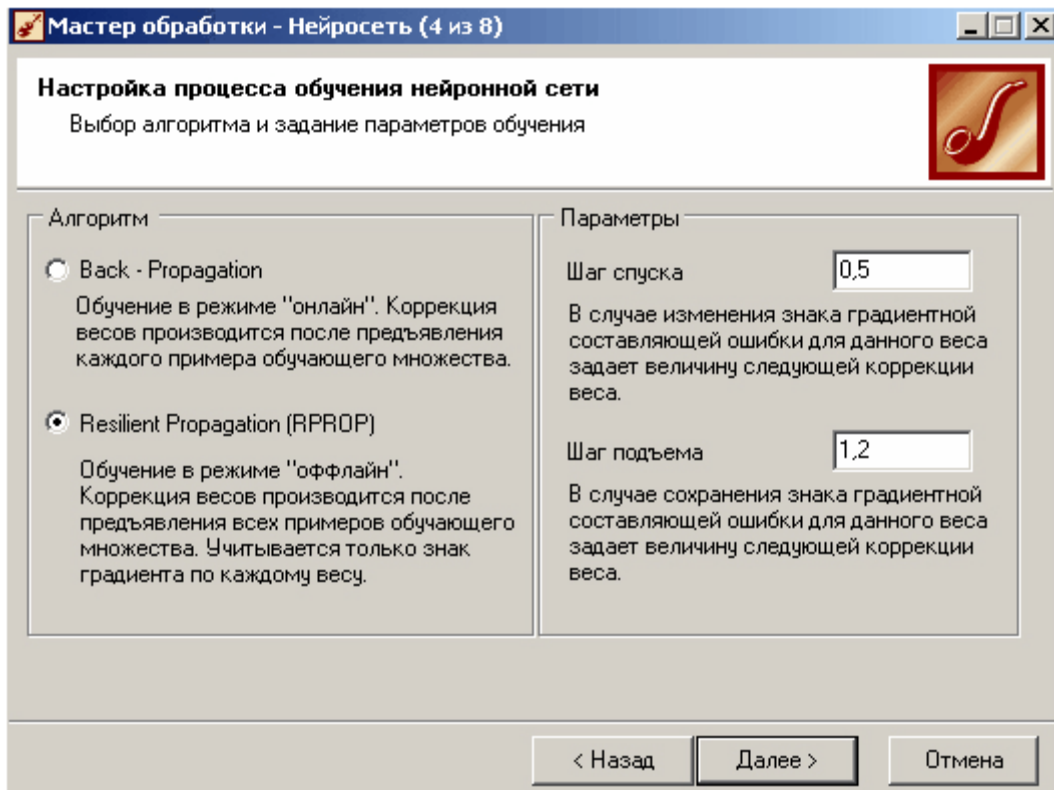
На следующем шаге укажем разбиение тестового и обучающего множеств.



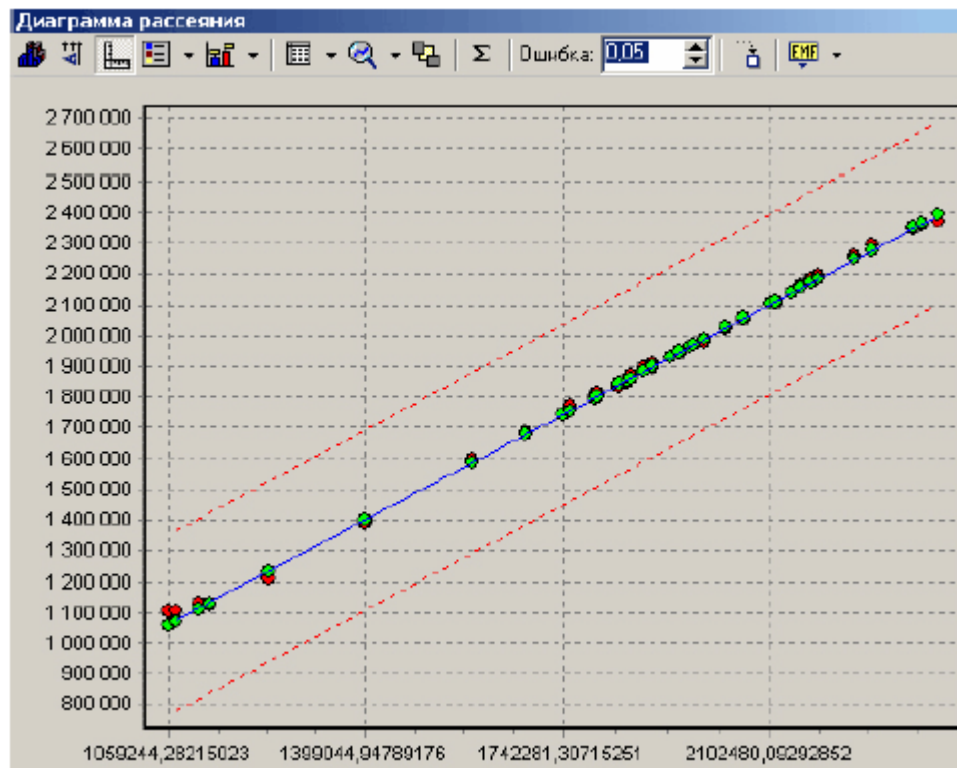
Перейдем к следующему шагу, на котором отметим необходимое количество слоев и нейронов в нейросети.



Перейдя далее, выберем алгоритм обучения нейросети.



После построения модели для просмотра качества обучения представим полученные данные в виде диаграммы и диаграммы рассеяния.



В Мастере настройки диаграммы выберем для отображения поля "Количество" и "Количество\_OUT" - реальное и спрогнозированное значение.

Метка столбца	Тип данных	Цвет
<input type="checkbox"/> Количество-6	9.0 Вещественный	<input type="checkbox"/>
<input type="checkbox"/> Количество-5	9.0 Вещественный	<input type="checkbox"/>
<input type="checkbox"/> Количество-4	9.0 Вещественный	<input type="checkbox"/>
<input type="checkbox"/> Количество-3	9.0 Вещественный	<input type="checkbox"/>
<input type="checkbox"/> Количество-2	9.0 Вещественный	<input type="checkbox"/>
<input type="checkbox"/> Количество-1	9.0 Вещественный	<input type="checkbox"/>
<input checked="" type="checkbox"/> Количество	9.0 Вещественный	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> Количество_OUT	9.0 Вещественный	<input checked="" type="checkbox"/>
<input type="checkbox"/> Количество_ERR	9.0 Вещественный	<input type="checkbox"/>

Тип:  Линии

Подписи по X: Дата (Год + Месяц)

Значения по X

Результатом будет два графика.

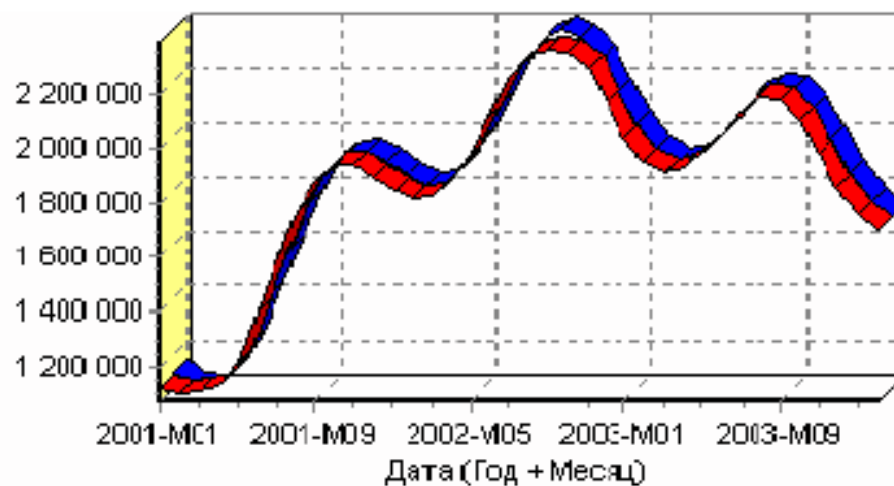
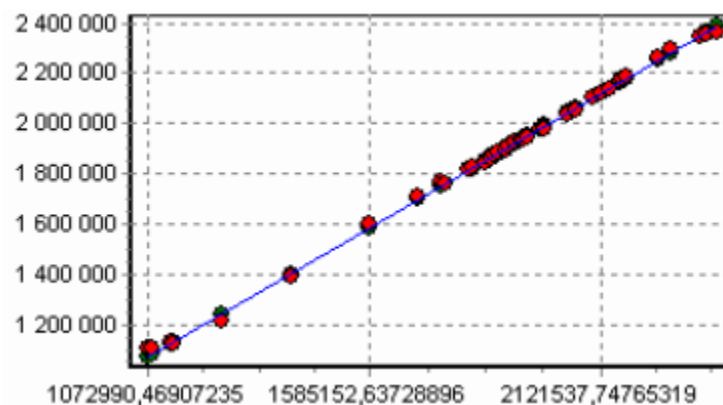


Диаграмма рассеяния более наглядно показывает качество обучения.

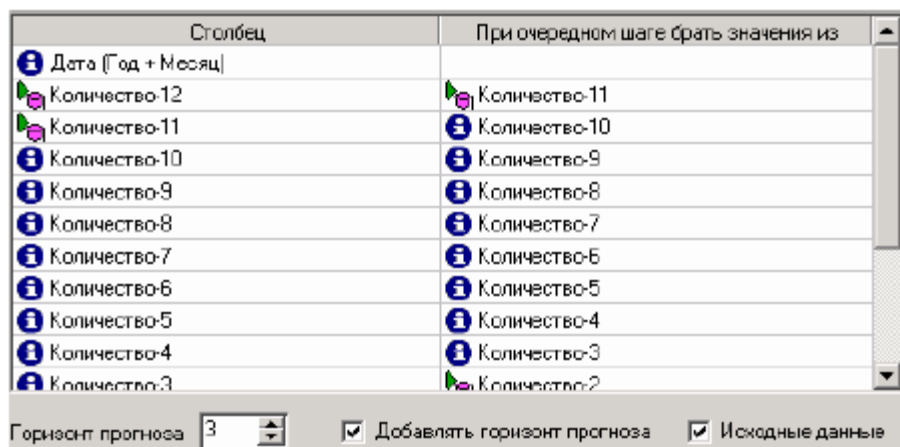


### Построение прогноза

Нейросеть обучена, осталось получить требуемый прогноз. Для этого открываем Мастер обработки и выбираем появившийся теперь обработчик "Прогнозирование".

На втором шаге Мастера предлагается настроить связи столбцов для прогнозирования временного ряда: откуда брать данные для столбца при очередном шаге прогноза. Мастер сам верно настроил все переходы, поэтому

остается только указать горизонт прогноза (на сколько вперед будем прогнозировать) равный трем, а также для наглядности следует добавить к прогнозу исходные данные, установив в Мастере соответствующий флажок.

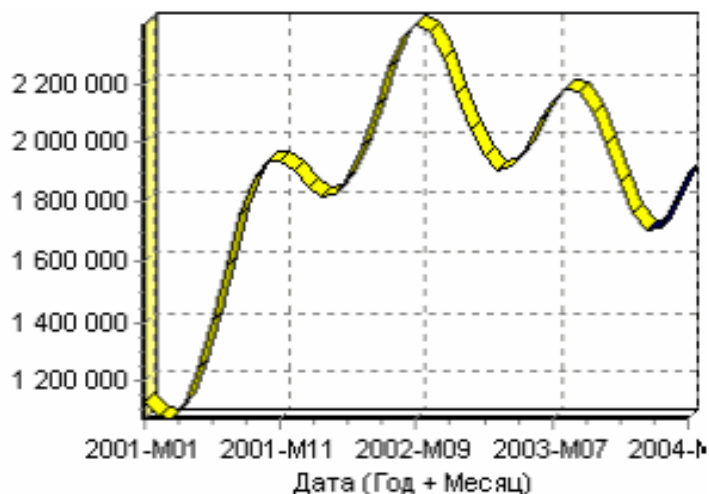


## Результат

После этого необходимо в качестве визуализатора выбрать "Диаграмму прогноза", которая появляется только после прогнозирования временного ряда.

В Мастере настройки столбцов диаграммы прогноза надо указать в качестве отображаемого столбец "Количество", а в качестве подписей по оси X указать столбец "Шаг прогноза".

Теперь аналитик может дать ответ на вопрос, какое Количество товаров будет продано в следующем месяце и даже два месяца спустя.



## Выводы

Данный пример показал, как с помощью **Deductor Studio** прогнозировать временной ряд.

При решении задачи были применены механизмы очистки данных от шумов, аномалий, которые обеспечили качество построения модели прогноза далее и соответственно достоверный результат самого прогнозирования количества продаж на три месяца вперед. Также был продемонстрирован



принцип прогнозирования временного ряда – импорт, выявление сезонности, очистка, сглаживание, построение модели прогноза и собственно построение прогноза временного ряда, а также экспорт результатов во внешний файл.

Подобный сценарий – основа любого прогнозирования временного ряда с той разницей, что для каждого случая приходится, как получать необходимый временной ряд посредством инструментов **Deductor Studio** (например, группировки), так и подбирать параметры очистки данных и параметры модели прогноза (например, структуры сети, если используется обучение нейронной сети, определение значимых входных факторов). В данном случае приемлемые результаты получились с настройками по умолчанию, в большинстве же случаев предстоит работа по их подбору (например, оценивая качество модели по диаграмме рассеяния).