


ПЕРВОЕ ВЫСШЕЕ ТЕХНИЧЕСКОЕ УЧЕБНОЕ ЗАВЕДЕНИЕ РОССИИ



**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
федеральное государственное бюджетное образовательное учреждение
высшего образования
САНКТ-ПЕТЕРБУРГСКИЙ ГОРНЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ


Руководитель программы
аспирантуры
доцент Ю.В. Ильюшин

**МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ САМОСТОЯТЕЛЬНОГО
ИЗУЧЕНИЯ ДИСЦИПЛИНЫ
ТЕХНОЛОГИИ ОБРАБОТКИ ИНФОРМАЦИИ**

Подготовка научных и научно-педагогических кадров в аспирантуре

Область науки:	2. Технические науки
Группа научных специальностей:	2.3. Информационные технологии и телекоммуникации
Научная специальность:	2.3.1. Системный анализ, управление и обработка информации, статистика
Отрасли науки:	Технические
Форма освоения программы аспирантуры:	Очная
Срок освоения программы аспирантуры:	3 года
Составитель:	д.т.н., проф. Трофимец В.Я.

Санкт-Петербург

Дисциплина «Технологии обработки информации» предназначена для аспирантов, специализирующихся в области проектирования и применения информационных систем и технологий в различных видах деятельности.

Цель дисциплины – овладение основными положениями методологии интеллектуального анализа данных (ИАД), развитие умений и навыков применения инструментальных средств ИАД в процессе обоснования управленческих решений.

Основными задачами изучения дисциплины являются:

- изучение основных теоретических положений анализа данных в контексте его применения для решения профессионально-ориентированных задач бизнес-аналитики;
- формирование умений и навыков, позволяющих эффективно применять методы и инструментальные средства ИАД в процессе обоснования управленческих решений.

Дисциплина «Технологии обработки информации» входит в составляющую «Дисциплины (модули), в том числе элективные, факультативные дисциплины (модули), дисциплины, направленные на подготовку к сдаче кандидатских экзаменов» образовательного компонента программы подготовки научных и научно-педагогических кадров в аспирантуре по научной специальности 2.3.1. Системный анализ, управление и обработка информации, статистика и изучается в 3 семестре.

Самостоятельная работа аспирантов

Самостоятельная работа аспиранта включает:

- тематическую работу с рекомендованной научной литературой;
- самостоятельное изучение разделов дисциплины;
- исследовательскую работу, анализ научных публикаций по теме курса;
- подготовку к зачету.

Самостоятельная работа обучающихся направлена на углубление и закрепление знаний, полученных на лекциях, выработку навыков самостоятельного активного приобретения новых, дополнительных знаний, подготовку к предстоящим учебным занятиям и промежуточному контролю.

Самостоятельная работа аспирантов - планируемая учебная и научно-исследовательская работа аспирантов, выполняемая во внеаудиторное время по заданию и при методическом руководстве преподавателя.

Целью самостоятельной работы аспирантов является овладение фундаментальными и профессиональными знаниями и умениями по профилю будущей специальности.

Основные задачи самостоятельной работы аспирантов

- изучение теоретического курса, углубление и расширение теоретического курса, углубление и расширение теоретической подготовки в области правовой охраны интеллектуальной собственности;
- формирование самостоятельного мышления, способностей к саморазвитию и самореализации;
- закрепление полученных теоретических знаний и практических умений;
- использование материала, полученного в ходе самостоятельных занятий в процессе ознакомления с нормативной, справочной документацией и специальной литературой.

Основными формами самостоятельной работы аспирантов являются:

- работа с учебной/научной литературой и углубление знаний при решении практических задач;
- подготовка к зачету.

Требуется изучить следующие темы и ответить на основные вопросы:

Тема 1. Введение в анализ данных

1. Чем различаются аналитический и информационный подходы к моделированию?
2. Перечислите формы представления данных их типы и виды.
3. Охарактеризуйте типы шкал измерений.
4. Опишите типовую структуру информационно-аналитических систем.
5. Опишите основные этапы интеллектуального анализа данных.

Тема 2. Технологии многомерного хранения, представления и обработки данных

1. Назовите отличительные черты информационных систем, ориентированных на операционную (транзакционную) обработку данных (OLTP). Назовите отличительные черты информационных систем оперативного анализа данных (OLAP).
2. Назовите общие свойства хранилищ данных. Назовите основные принципы построения хранилищ данных.
3. Назовите особенности реляционного хранилища данных. Назовите особенности гибридного хранилища данных. Назовите особенности виртуального хранилища данных.
4. В чем состоит суть концепции организации многомерной модели данных? Опишите структуру многомерного хранилища данных.
5. Охарактеризуйте возможности OLAP-кубов.

Тема 3. Технологии предобработки и очистки данных

1. В чем состоит суть оценки качества данных по их происхождению.
2. Перечислите состав типового набора инструментов предобработки данных в аналитических приложениях.
3. В чем состоит суть фильтрации данных?
4. В чем состоит суть обработки дубликатов и противоречий?
5. Перечислите виды аномалий данных. Охарактеризуйте обнаружение аномальных значений специальными методами.

Тема 4. Технологии статистической обработки данных

1. В чем польза предварительного исследования данных в дополнение к результатам автоматического компьютерного анализа?
2. Почему доверительный интервал полезнее, чем оценка значения?
3. Какие проблемы возникают при наличии выбросов значений?
4. Что такое дерево вероятностей?
5. В чем разница между корреляционным анализом и регрессионным анализом данных?

Тема 5. Технологии Data Mining

1. Какие технологии относятся к интеллектуальному анализу данных?
2. Что такое обучение с учителем и обучение без учителя?
3. Для чего служат алгоритмы g-mean и k-mean?
4. В чем состоит суть алгоритма CART?
5. В чем состоит суть алгоритма Apriori?

6.2. Оценочные средства для проведения промежуточной аттестации

6.2.1. Примерный перечень вопросов/заданий для подготовки к диф. зачету:

1. Аналитический и информационный подходы к моделированию.
2. Структурированные данные: формы представления данных, типы данных, виды данных.
3. Основные этапы интеллектуального анализа данных.
4. Машинное обучение и классы задач Data Mining.
5. Структура и архитектура информационно-аналитических систем и систем поддержки принятия решений.
6. Информационные системы, ориентированные на операционную (транзакционную) обработку данных (OLTP). Информационные системы оперативного анализа данных (OLAP).

7. Общие свойства хранилищ данных. Реляционные, гибридные и виртуальные хранилища данных. Витрины данных.
8. Концепция организации многомерной модели данных.
9. Стадии создания хранилищ данных. Аналитические платформы построения хранилищ данных.
10. Технологии реализации хранилищ данных компаний Microsoft и Oracle.
11. Структура многомерного хранилища данных, основные возможности OLAP-кубов.
12. Организация облачных хранилищ данных.
13. Уровни очистки данных. Оценка пригодности данных к анализу.
14. Оценка качества данных по их происхождению.
15. Предобработка данных и ее отличие от очистки. Типичный набор инструментов предобработки в аналитическом приложении.
16. Фильтрация данных. Обработка дубликатов и противоречий
17. Виды аномалий данных. Обнаружение аномальных значений специальными методами.
18. Трансформация, объединение и квантование данных.
19. Дисперсионный анализ.
20. Ковариация и корреляция.
21. Простая и множественная линейная регрессия. Оценка соответствия линейной регрессии реальным данным.
22. Регрессия с категориальными входными переменными.
23. Множественная логистическая регрессия.
24. Простой байесовский классификатор.
25. Временной ряд и его компоненты. Трендовые модели прогнозирования.
26. Скользящее среднее и экспоненциальное сглаживание.
27. Ассоциативные правила. Алгоритм Apriori.
28. Методы поиска логических закономерностей.
29. Задачи кластерного анализа. Иерархические и итеративные методы кластеризации. Особенности кластеризации в качественных и количественных шкалах.
30. Кластеризация данных по матрице объект-признак. Кластеризация данных по матрице связи.
31. Алгоритм кластеризации *k*-means.
32. Сети и карты Кохонена.
33. Назначение компонентного и факторного анализа. Применение компонентного и факторного анализа к задачам ИАД.
34. Методы распознавания образов с учителем и без учителя.
35. Алгоритмы построения деревьев решений.
36. Информационный подход к моделированию нейрона Принципы построения нейронных сетей. Место нейронных сетей среди других методов решения задач ИАД.
37. Алгоритмы обучения нейронных сетей.
38. Алгоритм обратного распространения ошибки.
39. Особенности структуры нейронных сетей и ее влияние на свойства сети.
40. Ансамбли моделей. Бэггинг и бустинг.

ОТРАБОТКА ПРАКТИЧЕСКИХ ЗАДАНИЙ

Практическая работа 1. Импорт данных

Импорт данных является отправной точкой любого сценария анализа данных. Импорт в Deductor может осуществляться из таких популярных форматов хранения данных, как Excel, Access, MS SQL, Oracle, текстовый файл, хранилища данных и прочее. Кроме того, обеспечивается доступ к любому источнику данных посредством механизмов ADO или ODBC.

В демонстрационных примерах импорт данных осуществляется из:

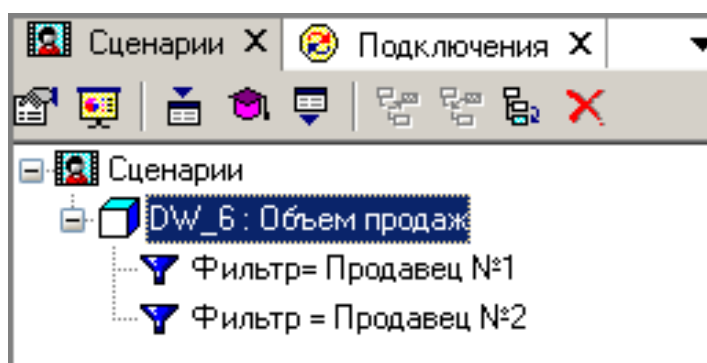
1. Импорт из текстового файла.
2. Импорт из хранилища данных.
3. Импорт из 1С: Предприятия.
- 4.

Импорт из хранилища данных

Deductor Warehouse 6 — многомерное хранилище данных, аккумулирующее всю необходимую для анализа предметной области информацию. Использование единого хранилища позволяет обеспечить непротиворечивость данных и централизованное хранение, а также автоматическую поддержку процесса анализа данных.

При работе с Deductor Warehouse от пользователя не требуется знание структуры хранения данных и языка запросов. Пользователь оперирует привычными терминами бизнес-среды — отгрузка, товар, клиент и т.п.

Данный пример импорта реализован в проекте "Демопример работы с хранилищем.ded" — выделенный фрагмент сценария.

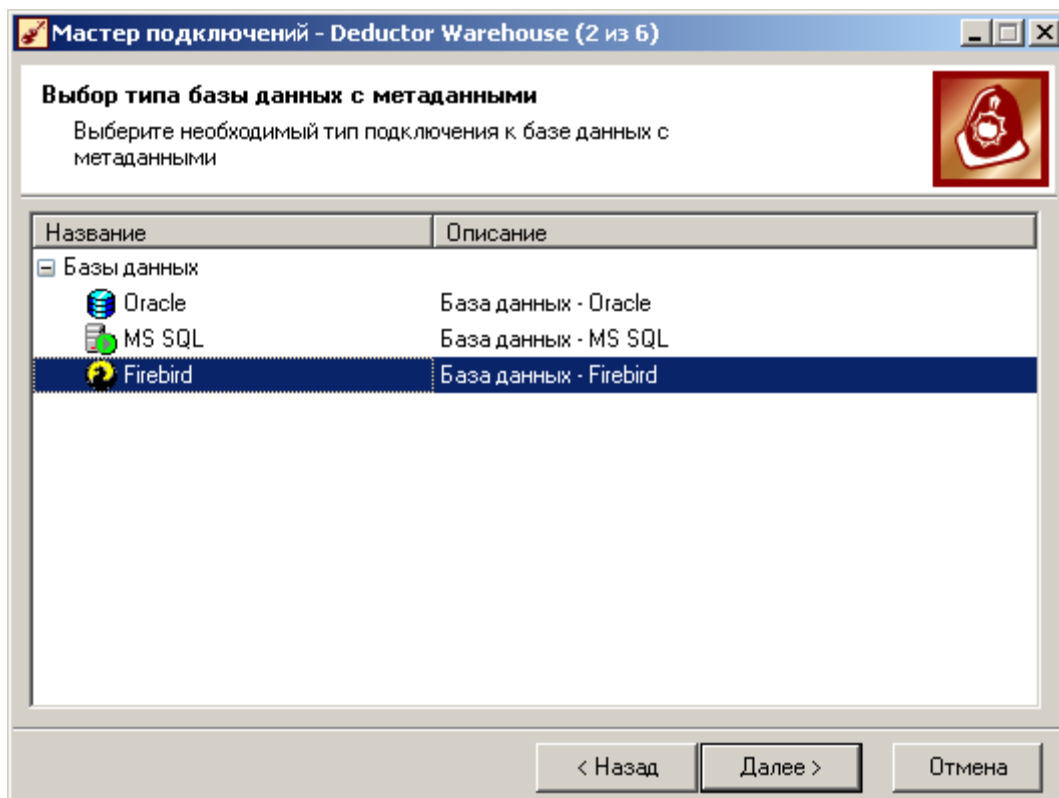


1. Настройка подключения хранилища данных

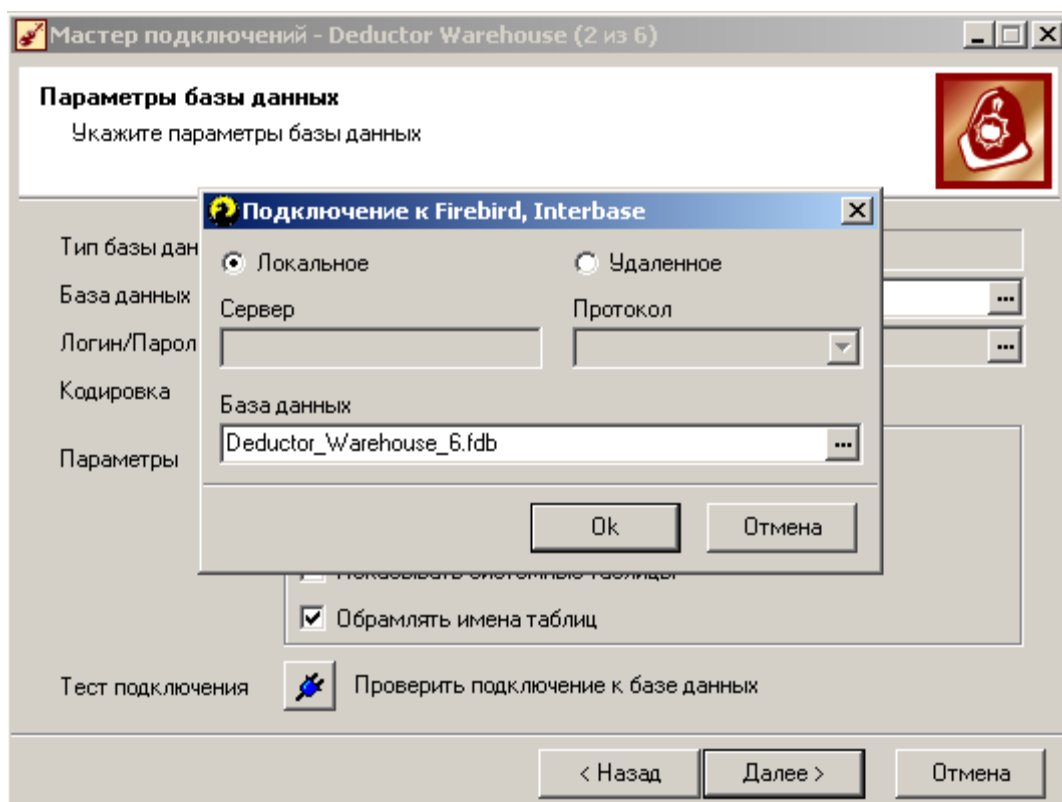
Для установления соединения с хранилищем данных воспользуемся Мастером настройки подключения на панели Подключения. Из списка подключений выбираем Deductor Warehouse.



На следующем шаге происходит выбор необходимой базы данных, выберем Firebird.

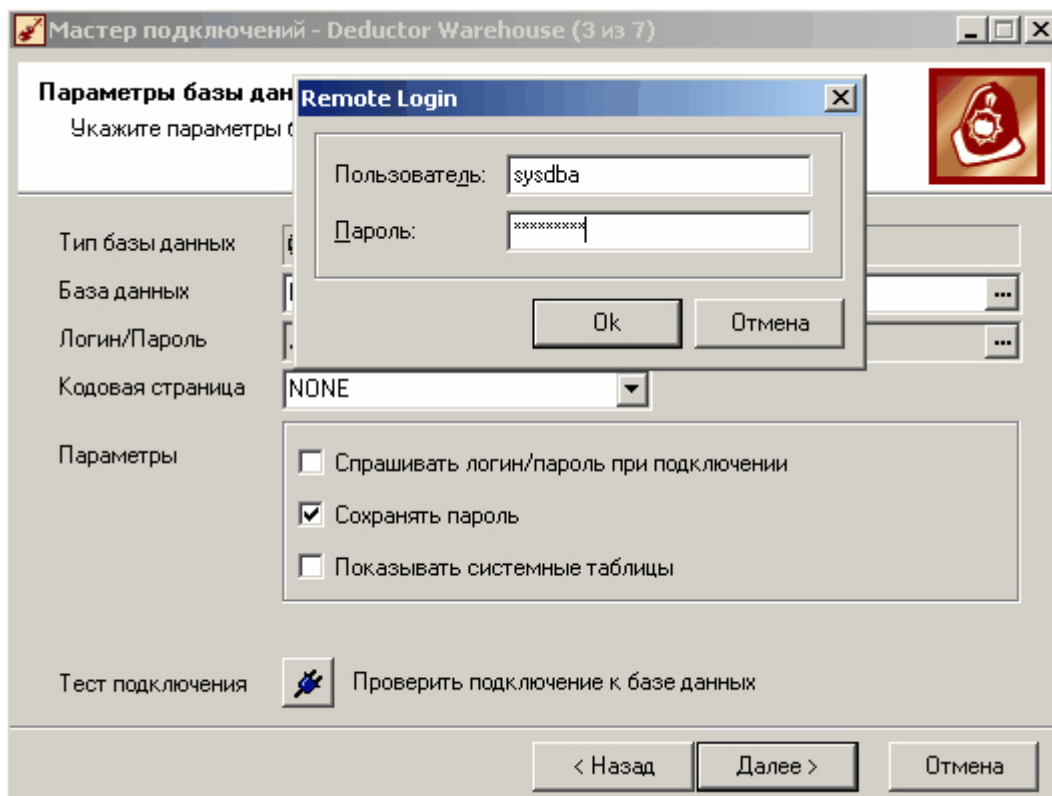


На третьем шаге необходимо задать параметры базы данных. В поле базы данных указываем нужную базу данных Deductor_Warehouse_6.fdb и локальное размещение файла. Файл Deductor_Warehouse_6.fdb входит в состав дистрибутива Deductor.



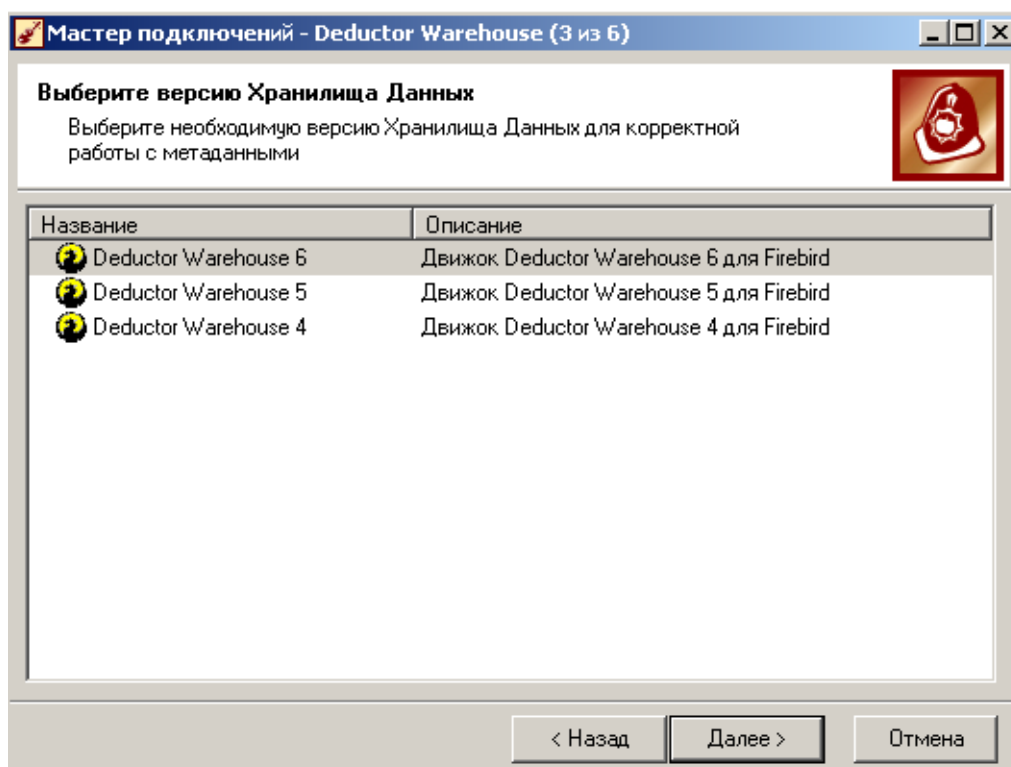
В поле Логин/Пароль укажем пользователя и пароль для доступа к базе данных. По умолчанию в базе данных Firebird имеется единственный пользователь sysdba с паролем

masterkey. Также необходимо включить опцию "Сохранять пароль", указав это галочкой в соответствующем поле. При необходимости можно проверить корректность подключения.

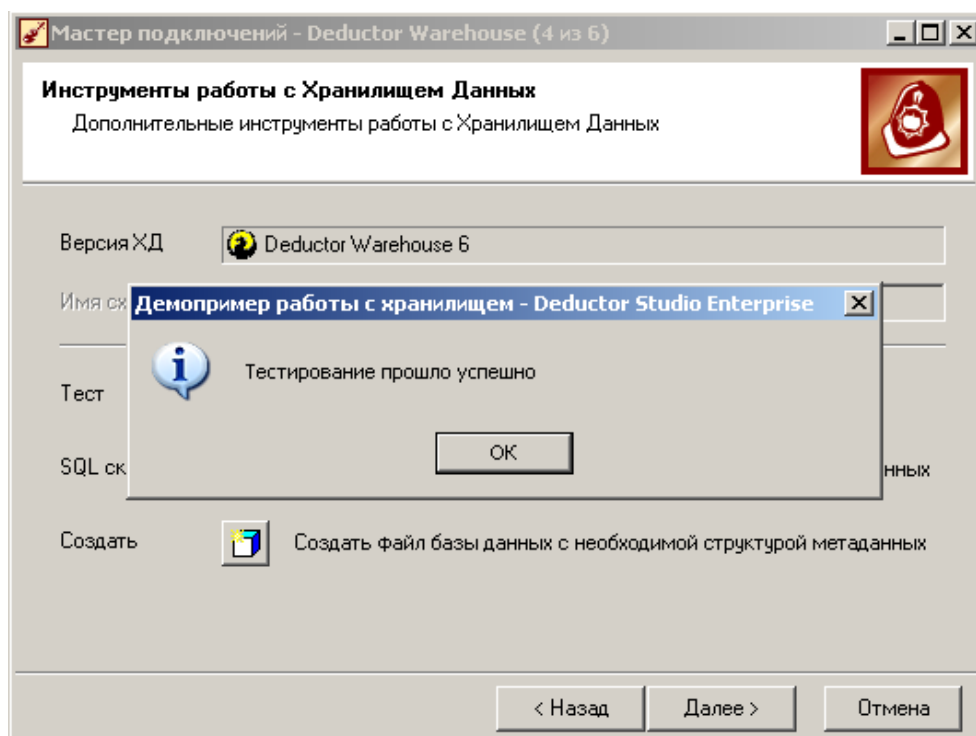


Далее необходимо указать версию хранилища данных.

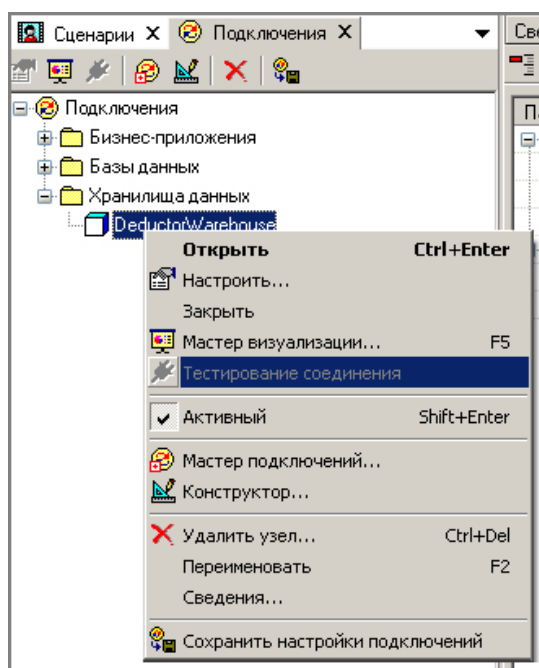
Deductor поддерживает хранилища данных трех версий Deductor Warehouse 4, Deductor Warehouse 5 и Deductor Warehouse 6. Для данного демопримера выбираем Deductor Warehouse 6.



На следующем шаге Deductor позволяет сгенерировать SQL скрипт для создания хранилища, создать требуемую структуру для метаданных, а также проверить наличие необходимой структуры метаданных в СУБД. Проверим правильность ввода информации о хранилище, нажав на Тест. Если тестирование прошло успешно, переходим на следующий этап.



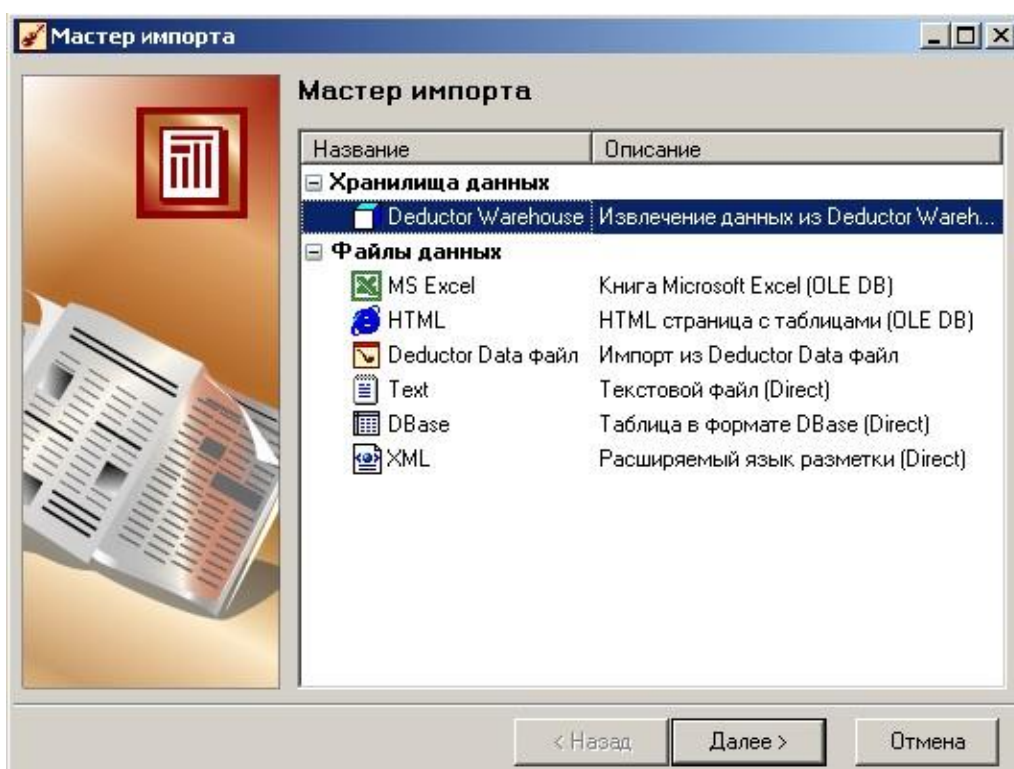
На следующих шагах Мастера подключения указывается способ отображения информации о подключении и метка хранилища (оставим эти параметры по умолчанию). В результате во вкладке подключения появится новое хранилище данных. Теперь хранилище подключено и готово к использованию.



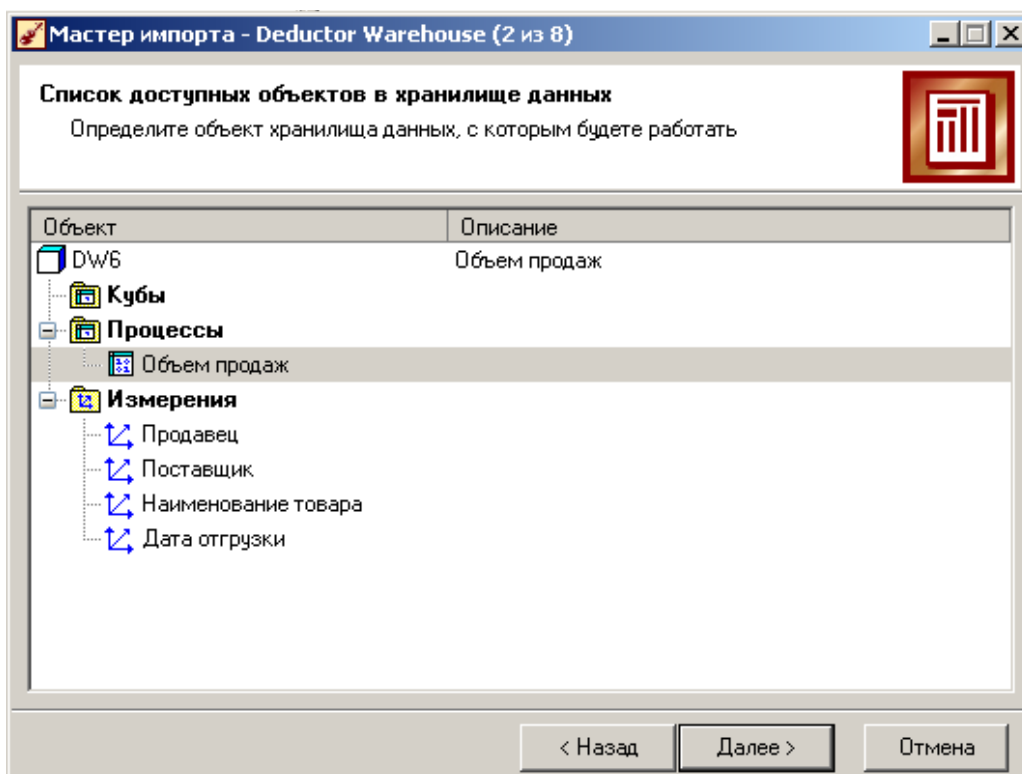
2. Импорт из Хранилища данных

На панели Сценарии воспользуемся Мастером импорта.

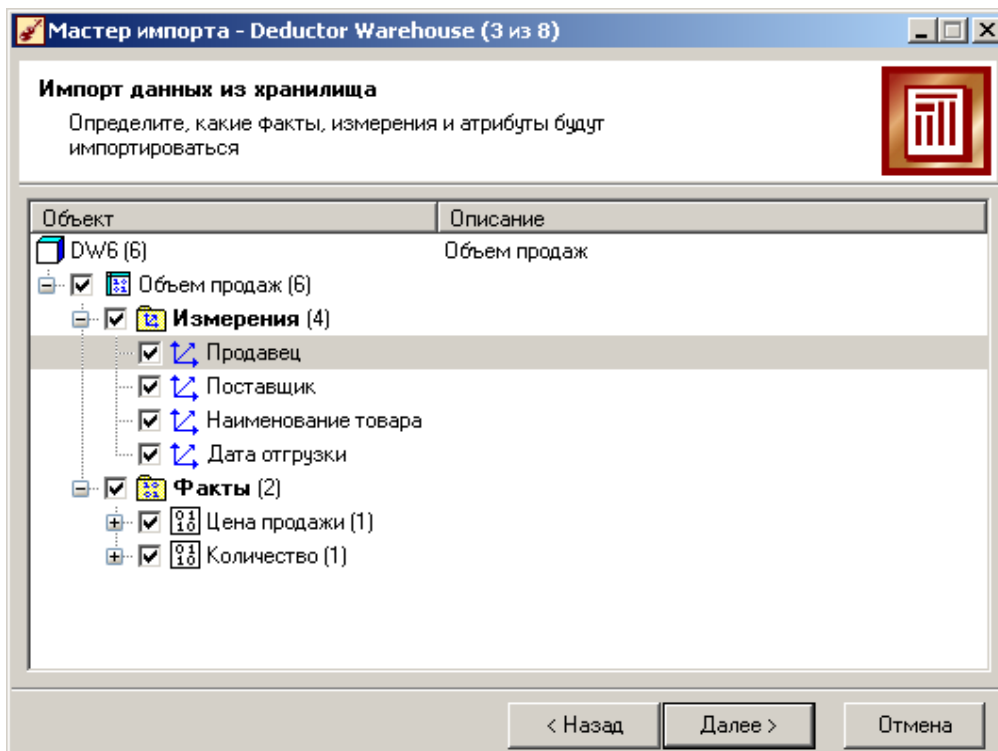
На первом шаге импорта выбираем Deductor Warehouse, далее укажем необходимое хранилище данных, в котором хранится интересующая информация.



После выбора хранилища следует окно выбора объекта хранилища данных. В данном окне необходимо указать куб, измерение или процесс, из которого будет произведен импорт. Укажем процесс "Объем продаж".

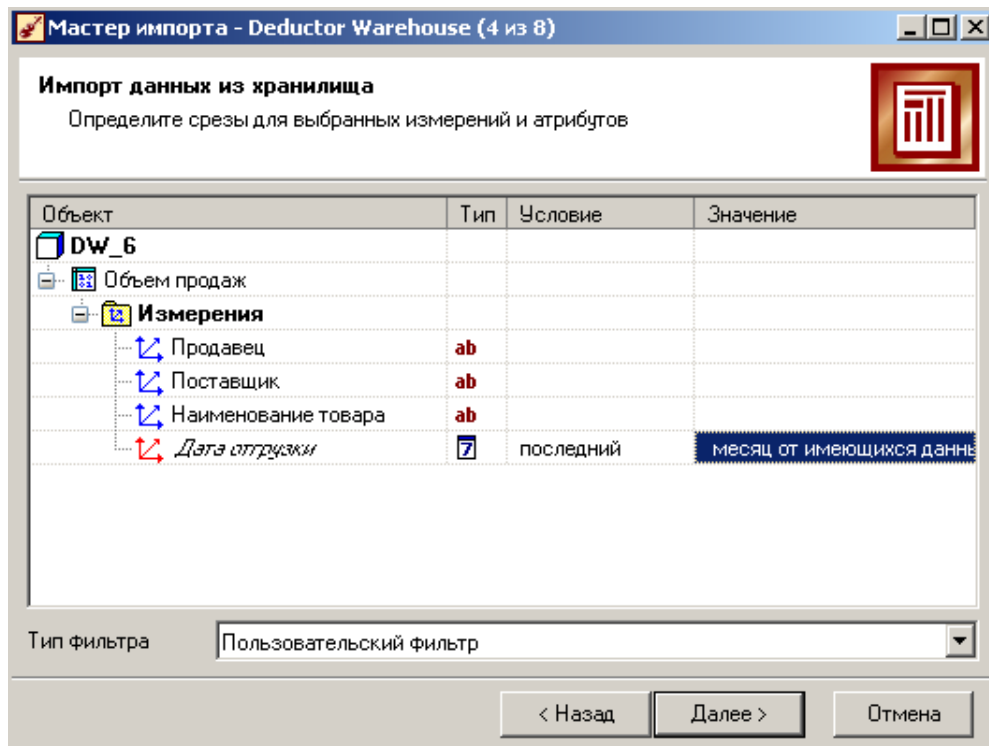


После выбора процесса появляется окно, где можно отобразить необходимые измерения и факты для импорта. Помечаем все указанные факты и измерения.

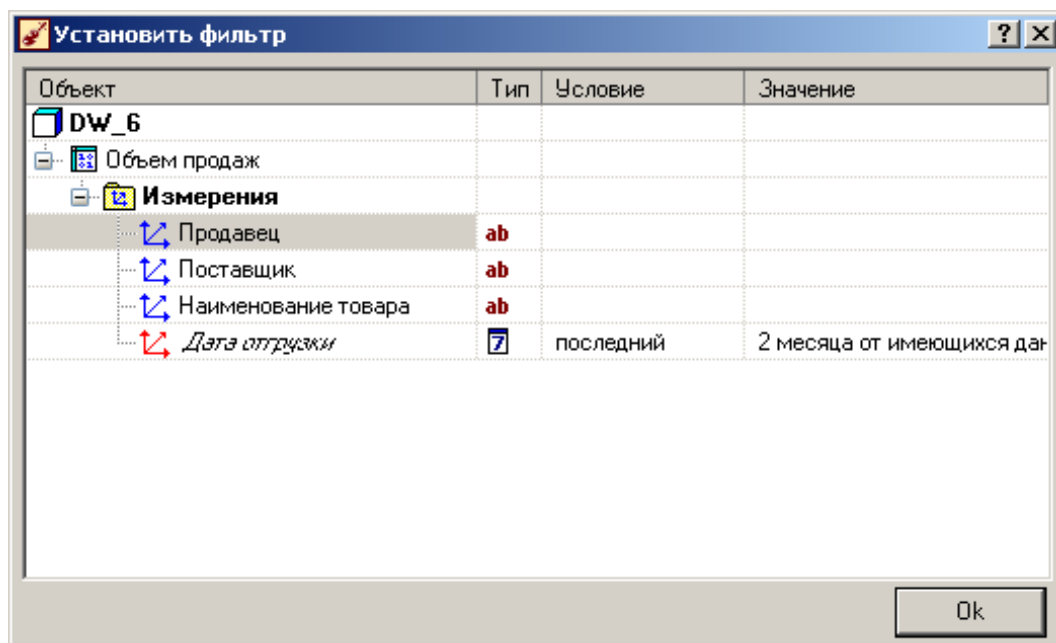


В следующем окне следует указать срезы для импортируемых измерений и установить тип фильтра. Всего фильтров три: статический для задания постоянного диапазона импорта данных; пользовательский для формирования диапазона импорта по требованию аналитика при каждой загрузке данных и переменный формирующий диапазон данных импорта в соответствии с переменными заложенными в проект.

Зададим пользовательский фильтр, который сформирует нам объем продаж за последний месяц, для этого установим для измерения "Дата отгрузки" условие "последний", а в поле месяц выберем "последний месяц от имеющихся данных".



Каждый раз при запуске импорта данных аналитик будет видеть окно, в котором он должен будет подтвердить или перенастроить условия импорта данных.



В следующем шаге происходит запуск процесса импорта при нажатии на кнопку Пуск.

3. Результат

После завершения процесса импорта нужно указать способ отображения. Выберем Куб. В результате получаем данные о продажах, готовые для дальнейшей обработки.

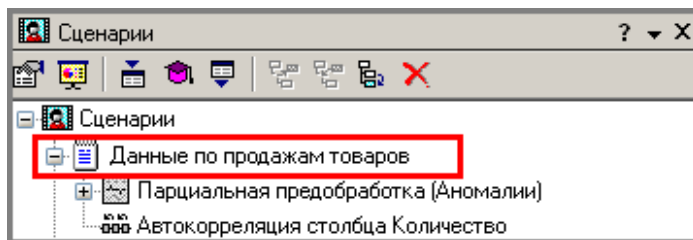
Куб						
Поставщик		Дата отгрузки		Продавец		
				Продавец №1		Продавец №2
				Продавец №2	Продавец №3	
Наименование товара	Σ Цена пр	Σ Количес	Σ Цена пр	Σ Количес	Σ Цена пр	Σ Количес
L-Тироксин 100 Берлин-Хеми таб. №100	100,10	3,00				
Аевит капс. 0.2 x10	12,00	50,00				
Азитромицин-АКОС капс. 250мг №6	115,60	4,00				
Аир корневища пачка упак. 50г					10,52	15,00
Актовегин амп. 2мл x25			700,00	50,00		
Аминазин амп. 2,5% 2мл №10	27,50	5,00				
Аминостерил КЕ 10% фл. 500мл	359,20	10,00				
Амиодарон (кордарон) таб. 200мг x30	53,00	20,00				
Амлодипин 0.005г №30	91,40	10,00				
Амлодипин 10мг №30	150,00	10,00				
Аммиака р-р 10% фл. 40мл						
Амоксиклав 1200 мг пор. д/в/в x5	813,80	10,00				
Анальгин 50% амп. 2мл x10	39,20	50,00				
Ардуан амп. 4мг 2мл x25	1 256,80	1,00				
Аскорутин таб. 0.33 x10	0,80	10,00				

Импорт из текстового файла

Рассмотрим пример импорта данных из текстового файла с разделителями, который будет необходим при апробировании технологий платформы Deductor на предлагаемых примерах.

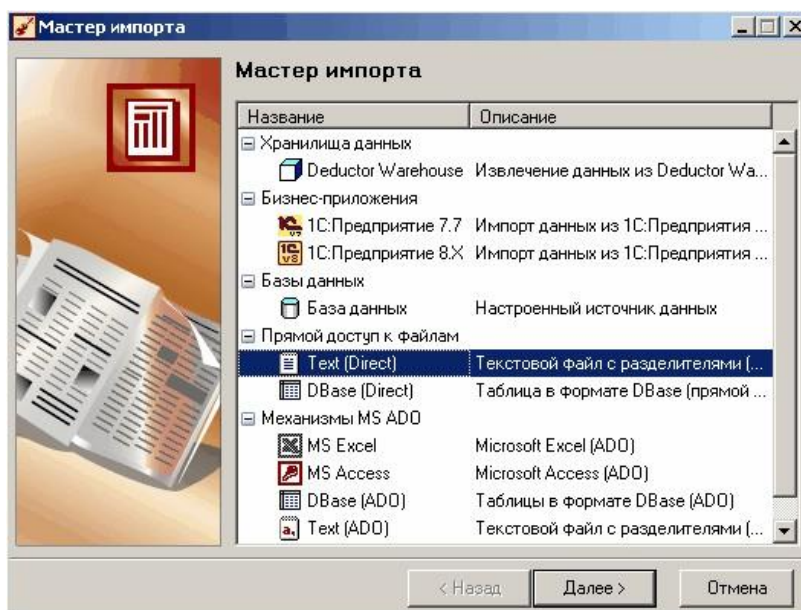
Покажем, как выполняется данный фрагмент сценария из проекта

"Демонстрация анализа данных.ded".

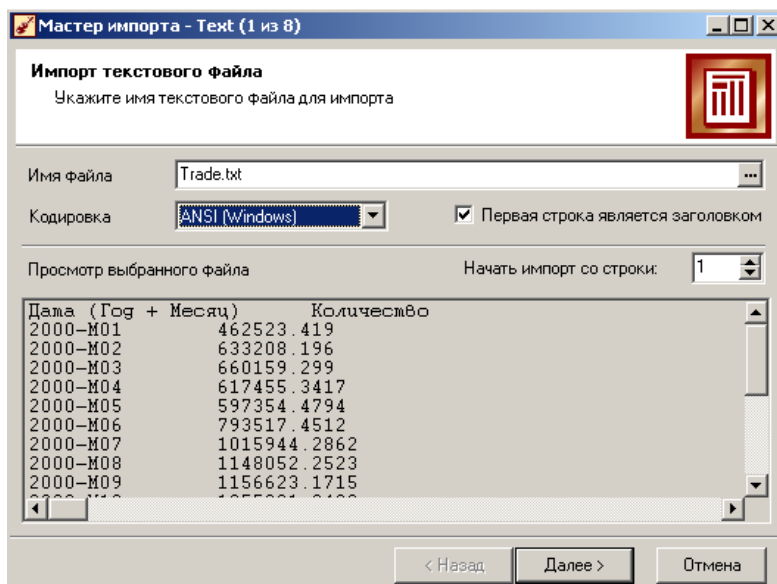


1. Импорт данных из файла

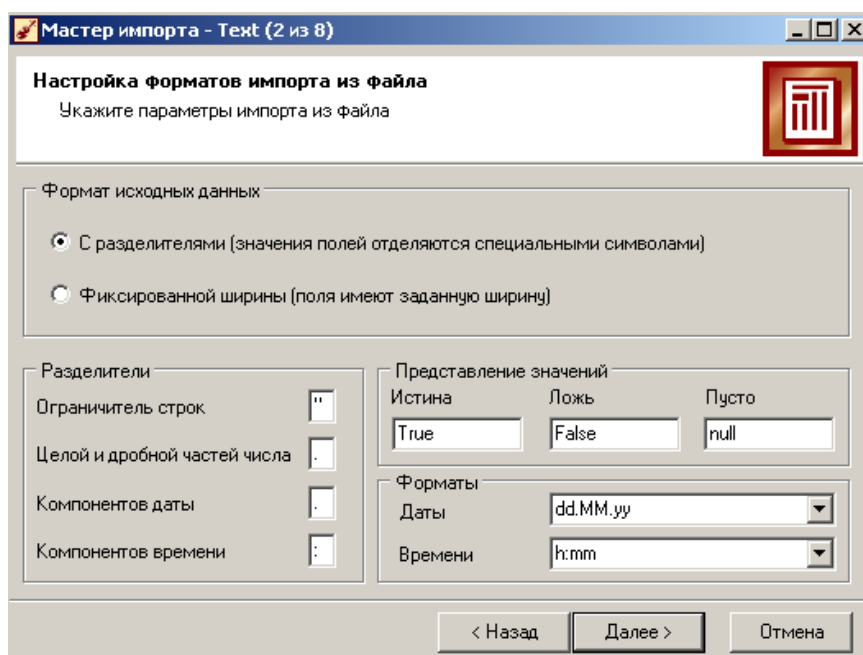
Импорт осуществляется путем вызова Мастера импорта на панели "Сценарии"



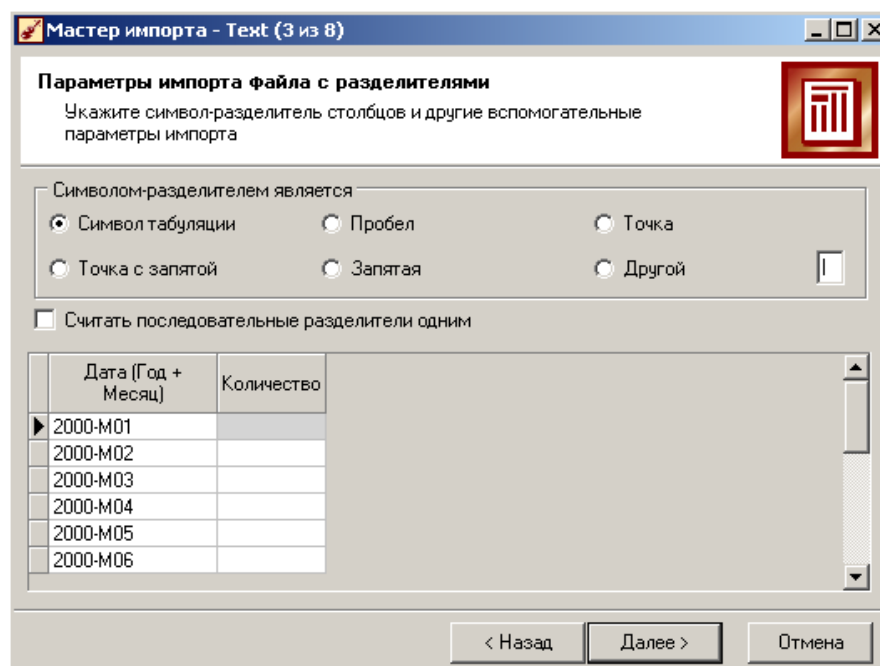
После запуска Мастера импорта укажем тип импорта "Текстовый файл с разделителями". Далее перейдем к настройке параметров импорта. Отметим имя файла, из которого необходимо получить данные (пример для парциальной обработки). В окне Мастера импорта текстового файла можно увидеть содержание данного файла, задать строку с которой будет осуществляться импорт, определить первую строку страницы как заголовок, а также при необходимости подобрать кодировку символов. Оставим текущую настройку без изменения.



На следующем шаге Мастера настраиваются параметры импорта данных из файла, задается формат разделителя исходных данных, ограничитель строк, разделитель целой и дробной частей вещественного числа, разделитель компонентов даты и ее формат.



В данном случае параметры по умолчанию на странице Мастера установлены верно и их оставляем без изменения, переходим к следующему шагу настройки разделителей.

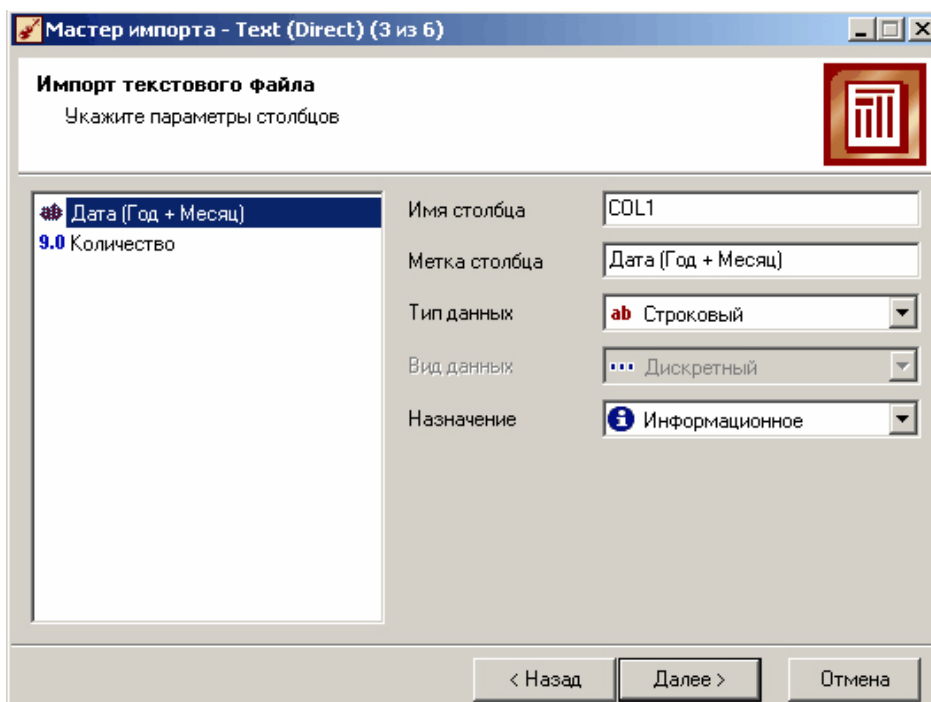


Проверим правильность автоматического определения разделителя столбцов, им является символ табуляции, оставим эту настройку так же без изменения.

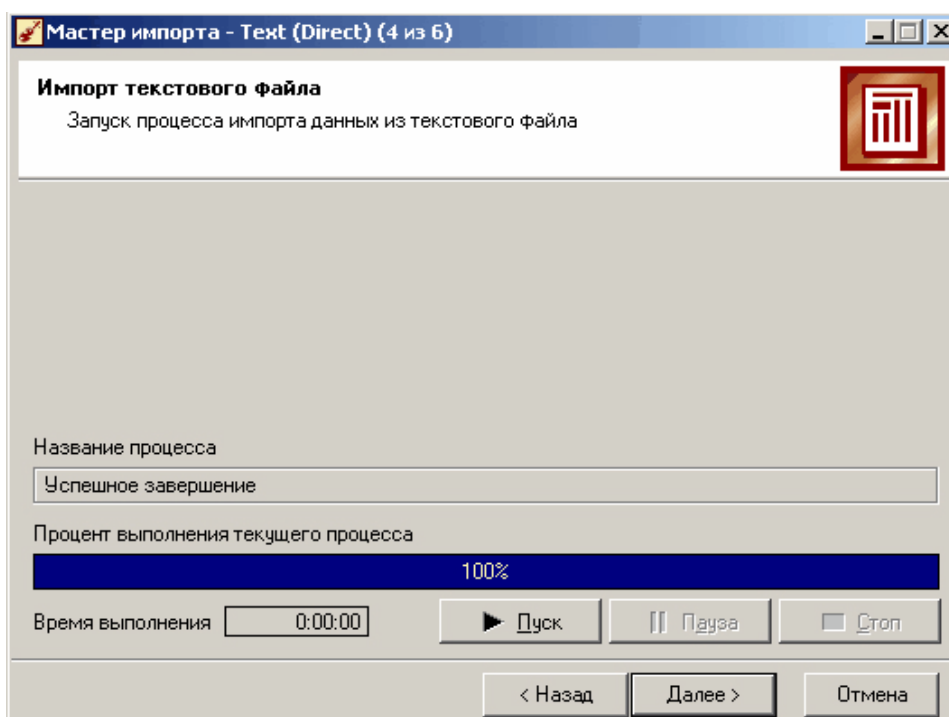
2. Настройка параметров столбцов

На этом шаге Мастера предоставляется возможность настроить параметры каждого поля: имя, метку, размер, тип данных, вид данных и назначение. Некоторые свойства (например, тип данных) можно задавать сразу для группы полей. Вид данных определяет, конеч-

ный ли это набор (дискретные) или бесконечный (непрерывные). Назначение столбцов выявляет характер их использования в алгоритмах обработки (при импорте можно оставить значение по умолчанию).

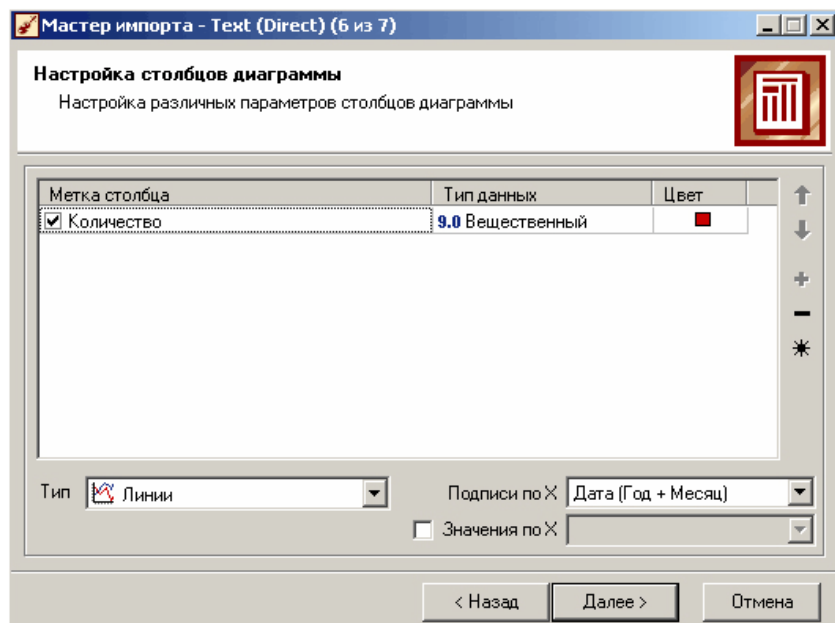


Указав параметры столбцов, запустим процесс импорта, нажав на кнопку "Пуск".

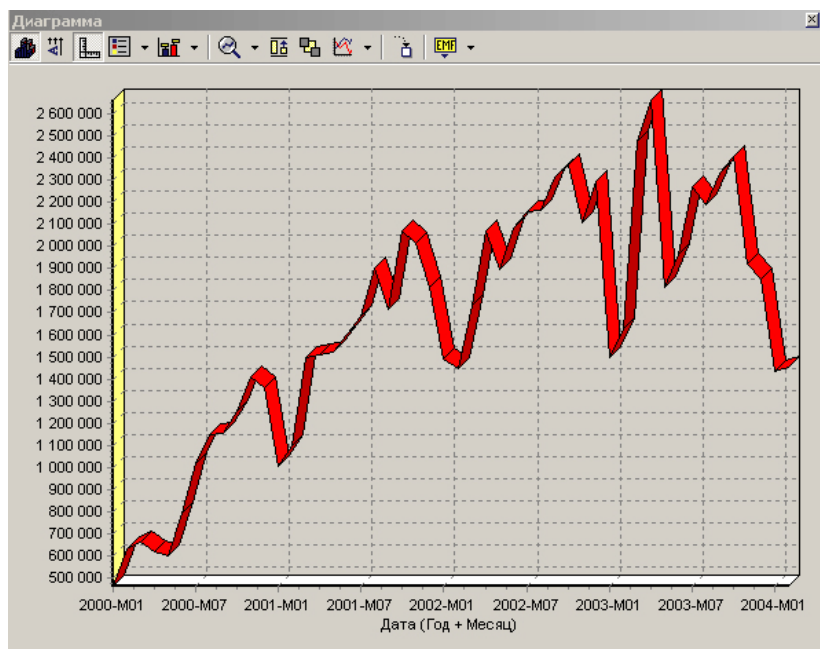


3. Результат

После импорта данных на следующем шаге Мастера необходимо выбрать способ отображения данных. В данном случае самым информативным является диаграмма, поэтому выберем ее. Параметры диаграммы настроим на следующем шаге Мастера импорта. На последнем этапе следует указать Метку(заголовок) импортированного файла.



После настройки отображения необходимых полей получаем требуемый результат.



Практическая работа 2. Многомерная визуализация

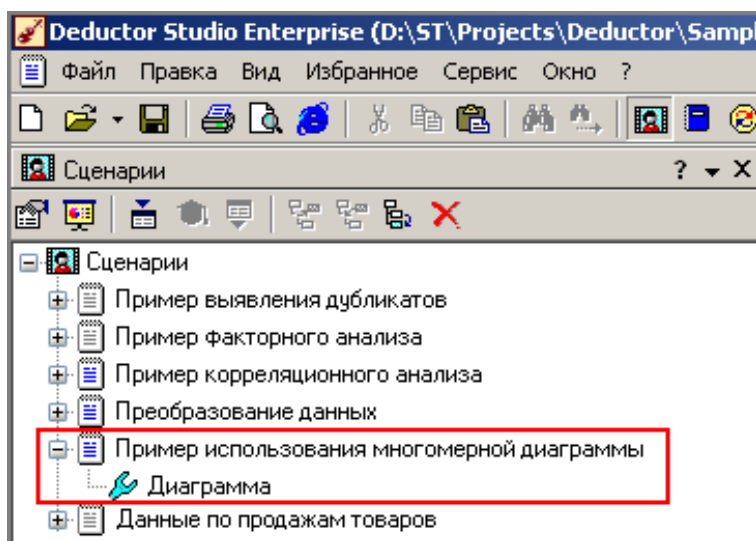
Многомерная визуализация является вспомогательным звеном в понимании анализируемой информации, позволяющим увеличить уровень восприятия данных и существенно облегчить процесс интерпретации полученных результатов. Применение многомерных механизмов визуализации способствует выявлению сложных аналитических зависимостей, представленных в графическом виде и определению общих тенденции в рассматриваемых процессах. В данном демопримере применяется метод многомерной диаграммы визуализации данных.

Многомерная диаграмма

Многомерная диаграмма является одним из самых наглядных визуализаторов, используемых при анализе данных. Она позволяет просмотреть трехмерную зависимость параметров рассматриваемых данных. Для ее построения в Deductor Studio используется визуализатор

"Многомерная диаграмма". Данный визуализатор имеет две настройки графического представления данных: поверхностный и топографический.

Использование этого визуализатора демонстрируется в выделенном фрагменте сценария проекта "Демоприменение анализа данных.ded".

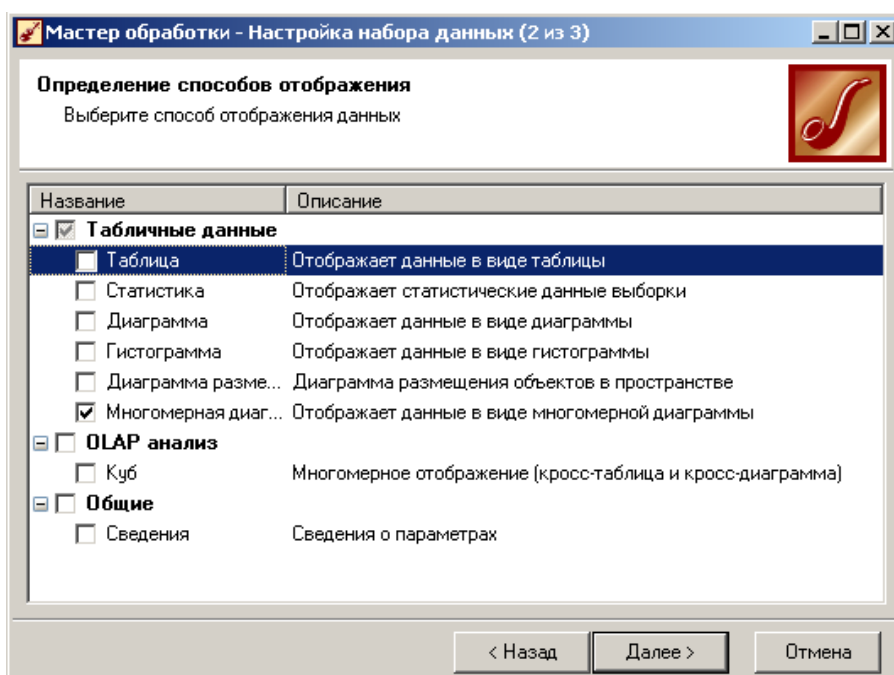


1. Исходные данные

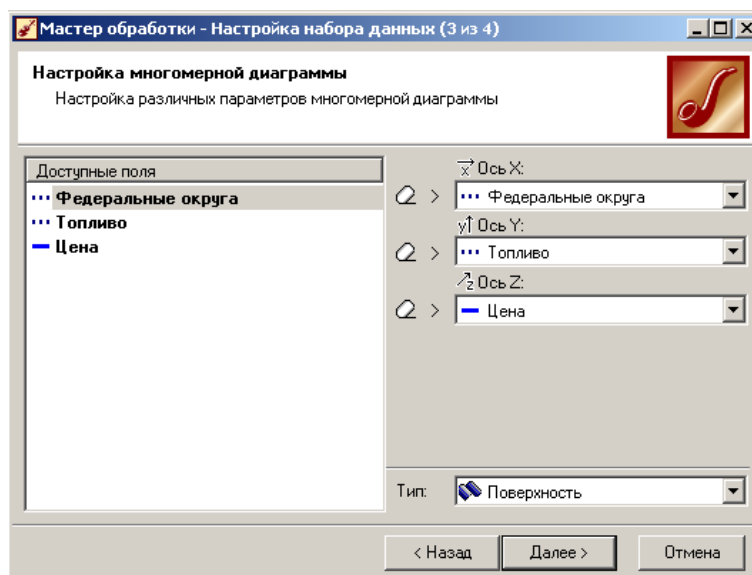
Рассмотрим построение многомерной диаграммы на примере данных из файла "fuel.txt". Он содержит таблицу с информацией о стоимости топлива в регионах РФ. Многомерная диаграмма позволяет увидеть различия в цене покупаемого топлива по регионам.

2. Выполнение настройки

Осуществим построение многомерной диаграммы в специально созданном для этого узле "Диаграмма", на основе обработчика "Настройка набора данных". Для этого выберем в настройках отображения данных рассматриваемого узла визуализатор "Многомерная диаграмма".

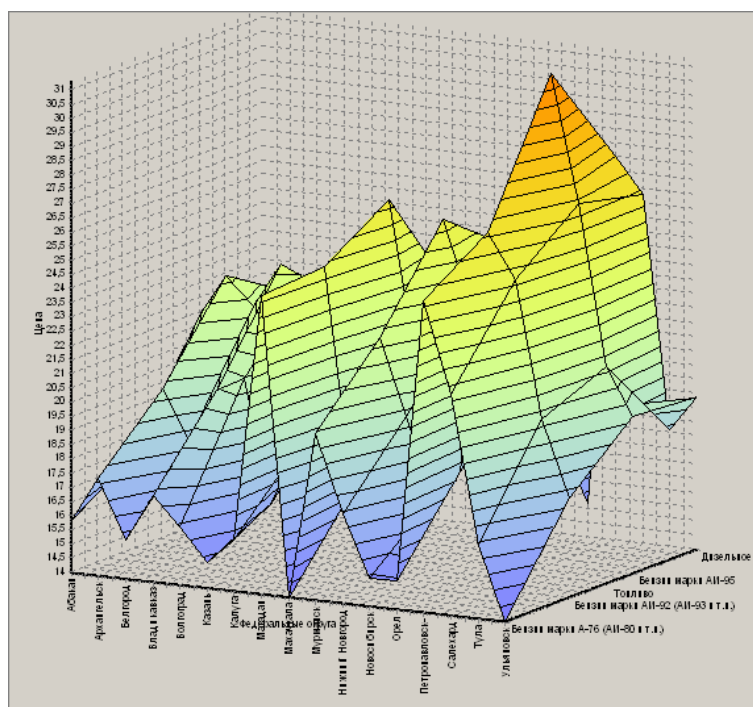


На следующем шаге определим оси координат. Выберем свойства на основе которых изменяется цена, в нашем случае, - это название регионов и наименование топлива. Обозначим координату x - "Федеральные округа", а у "Топливо", z - "Цена".



3. Результат

Выберем тип изображения диаграммы как "Поверхность", на котором будет показана объемная зависимость цены рассматриваемого топлива по регионам.



На диаграмме показаны перепады стоимости топлива в различных регионах не только графически, но и цветом. При необходимости можно просмотреть детализацию выбранных на графике точек, выбрав соответствующую настройку.

Практическая работа 3. Очистка данных

Часто исходные данные для анализа не годятся в связи с плохим качеством, так что вопрос подготовки данных для последующего анализа является очень важным. Обычно "сырые"

данные содержат в себе пропуски, различные шумы, за которыми трудно увидеть общую картину, а также аномалии – влияние случайно, либо редко происходивших событий. Очевидно, что влияние этих факторов на общую модель следует минимизировать, т.к. модель, учитывающая их, получится неадекватной. Минимизировать влияние шумов, аномалий и прочее можно, используя устойчивые к их воздействию алгоритмы анализа и применяя специализированные механизмы очистки.

В данном демопримере рассматриваются следующие методы очистки данных:

1. Парциальная обработка
2. Факторный анализ
3. Корреляционный анализ
4. Дубликаты и противоречия

Парциальная предобработка

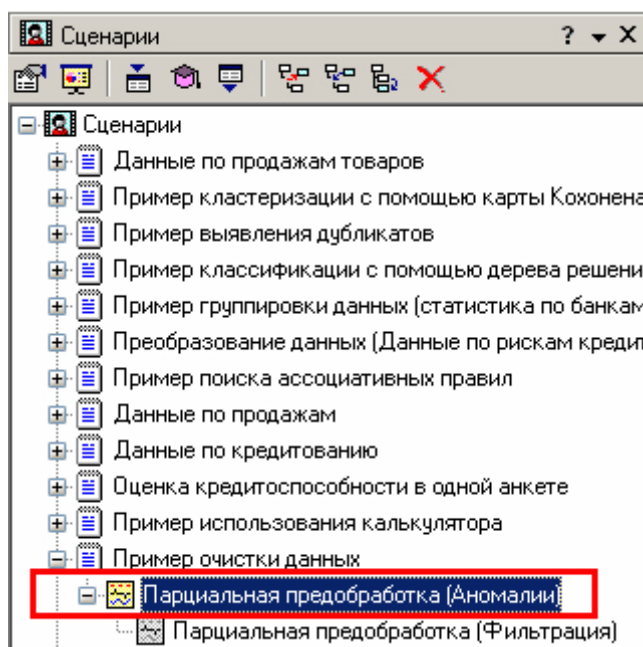
Парциальная предобработка служит для восстановления пропущенных данных, редактирования аномальных значений и спектральной обработки данных (например, сглаживания данных). Именно эти операции часто проводятся в первую очередь над данными.

1. Удаление аномалий
2. Спектральная обработка

Удаление аномалий

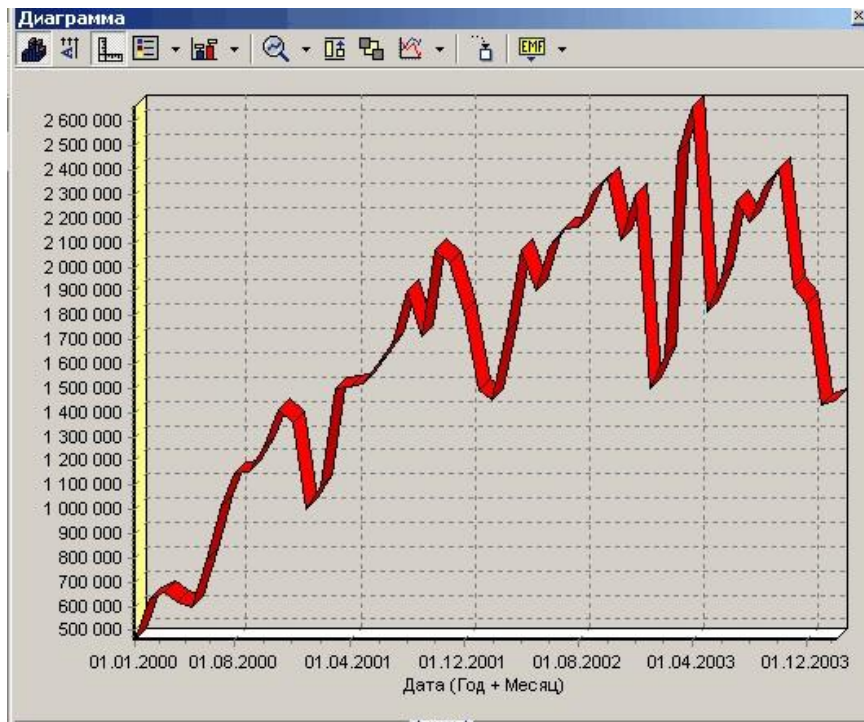
Присутствие аномалий при построении моделей оказывает на них большое влияние, ухудшая качество результата.

Покажем, как выполнялась парциальная обработка в выделенной части сценария проекта *"Демопример анализа данных.ded"*.



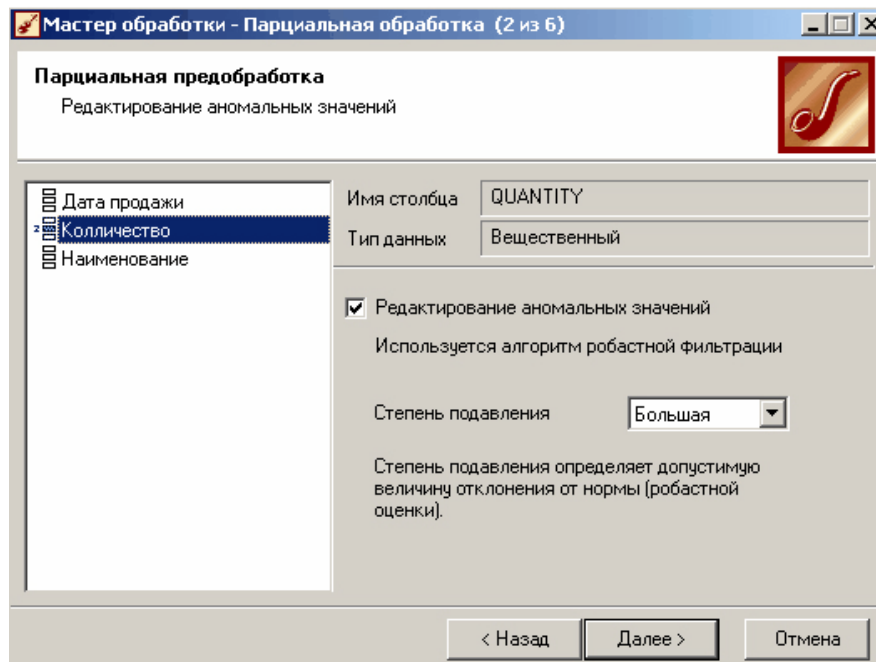
1. Исходные данные

В качестве примера возьмем данные из файла "Trade.txt". В данном файле находятся данные о продажах за некоторый период.



2. Обработка данных

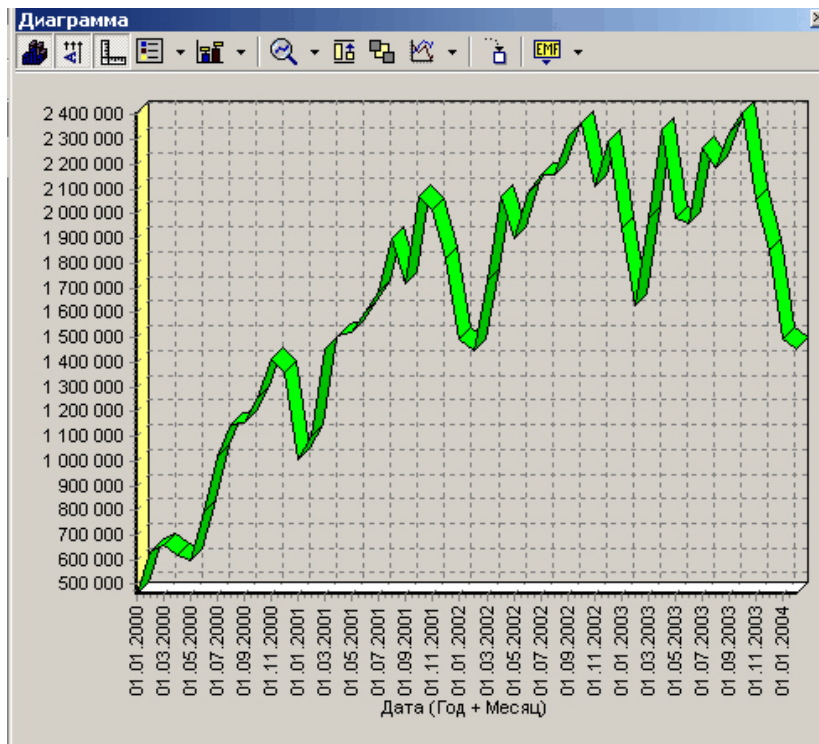
Как видно из диаграммы, выбросы ухудшают статистическую картину распределения данных. Воспользуемся Мастером обработки и выберем частичную обработку.



В Мастере парциальной предобработки на втором шаге выбираем поле "Количество" и указываем ему тип обработки "Редактирование аномальных значений", степень подавления "Большая". Так как больше никаких действий над данными не планировалось, то переходим на шаг запуска процесса обработки и нажимаем "Пуск".

3. Результат

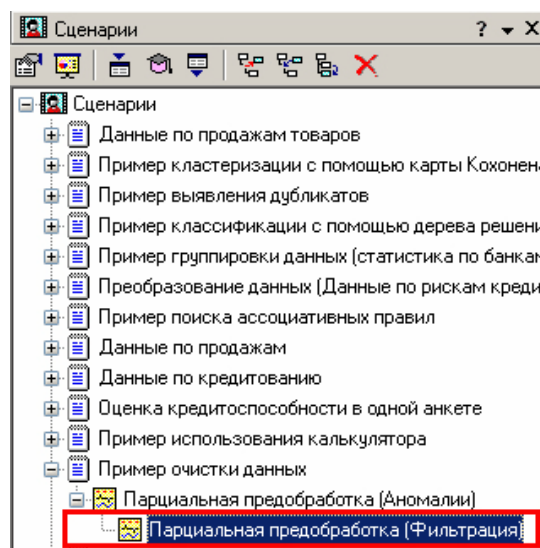
После выполнения процесса обработки на диаграмме видно, что выбросы уменьшились, стала проясняться реальная картина продаж.



Спектральная обработка

Сглаживание данных применяется для удаления шумов из исходного набора (что будет продемонстрировано позднее), а также для выделения тенденции, трудно обнаруживаемой в исходном наборе. Платформа Deductor Studio предлагает несколько видов спектральной обработки: сглаживание данных путем указания полосов пропускания, вычитание шума путем указания степени вычитания шума и вейвлета преобразования путем указания глубины разложения и порядка вейвлета.

Рассмотрим пример использования данного обработчика на выделенном фрагменте сценария проекта "Демонстример анализа данных.ded".

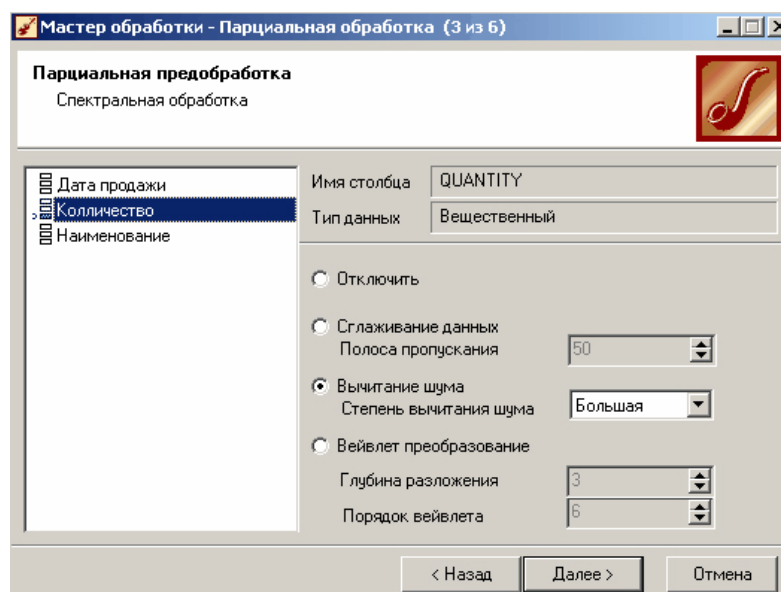


1. Исходные данные

Продemonстрируем такой метод спектральной обработки, как вычитание шума. Для этого продолжим работу с данными файла "Trade.txt".

2. Обработка данных

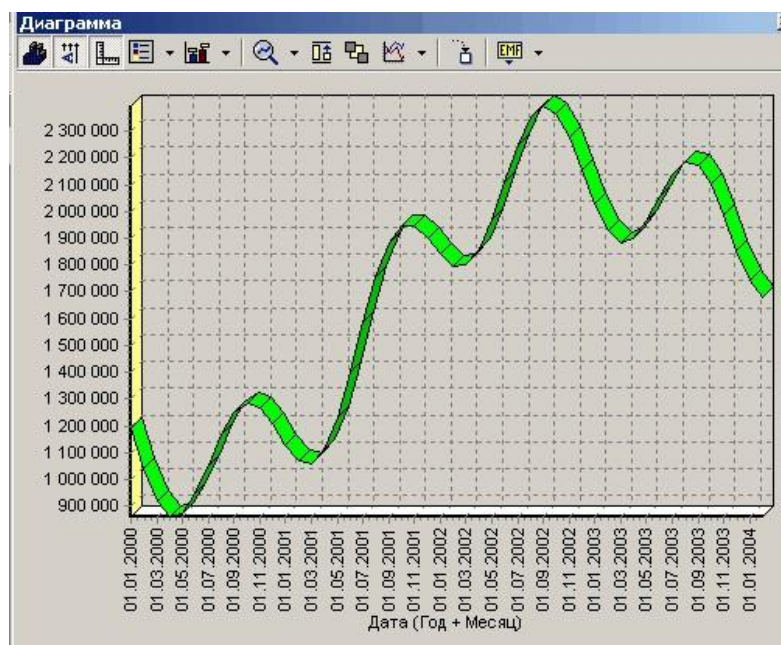
Сгладим данные при помощи парциальной обработки.



В Мастере парциальной предобработки на третьем шаге выбираем поле "Количество" и указываем ему тип обработки "Вычитание шума", степень подавления "Большая". Так как больше никаких обработок не планировалось, то переходим на шаг запуска процесса обработки и нажимаем "Пуск".

3. Результат

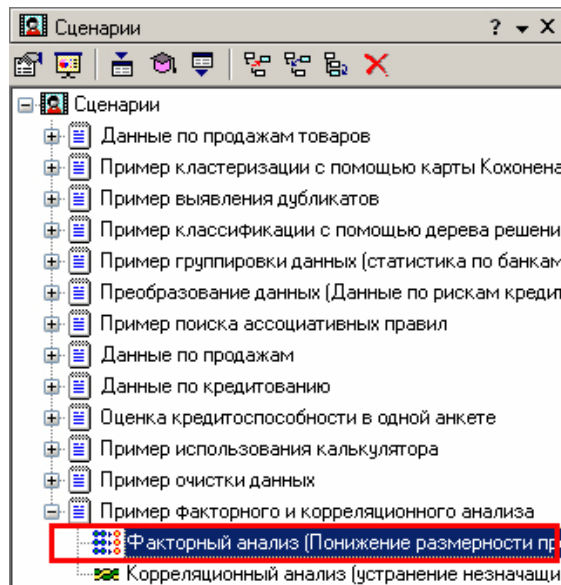
После выполнения процесса обработки выберем в качестве визуализации диаграмму.



Факторный анализ

Факторный анализ служит для понижения размерности пространства входных факторов. Обработку можно выполнять как в автоматическом режиме (с указанием порога значимости), так и вручную (основываясь на значениях матрицы значимости).

Покажем, как можно осуществить факторный анализ в данном фрагменте сценария проекта "Демонпример анализа данных.ded".



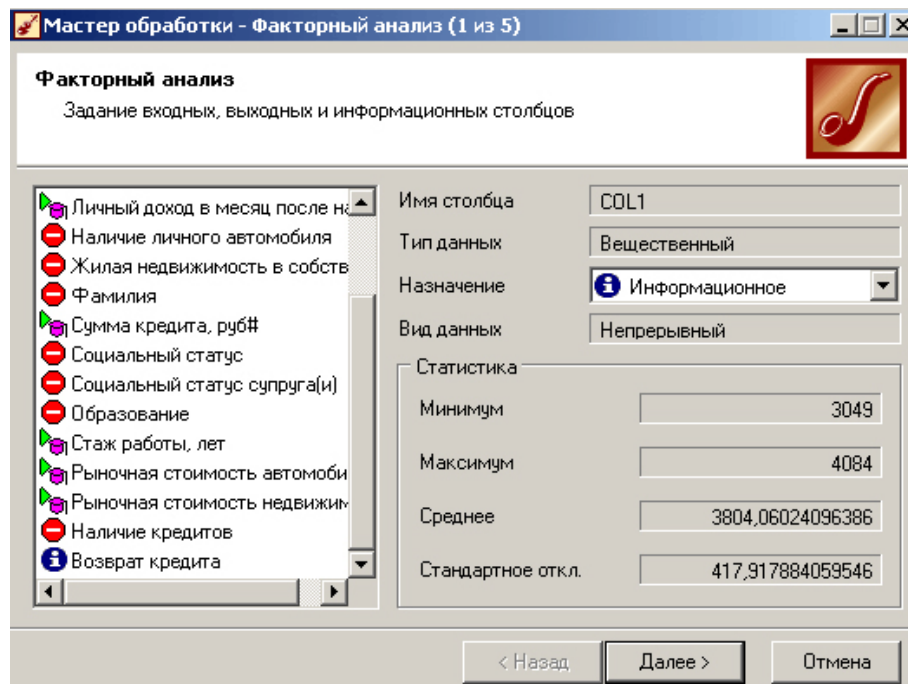
1. Исходные данные

Рассмотрим применение обработчика на примере данных из файла "Anketa1.txt". Он содержит таблицу с информацией о кредитах граждан. Попробуем выявить значение факторов, влияющих на возврат кредита.

2. Понижение размерности пространства входных факторов

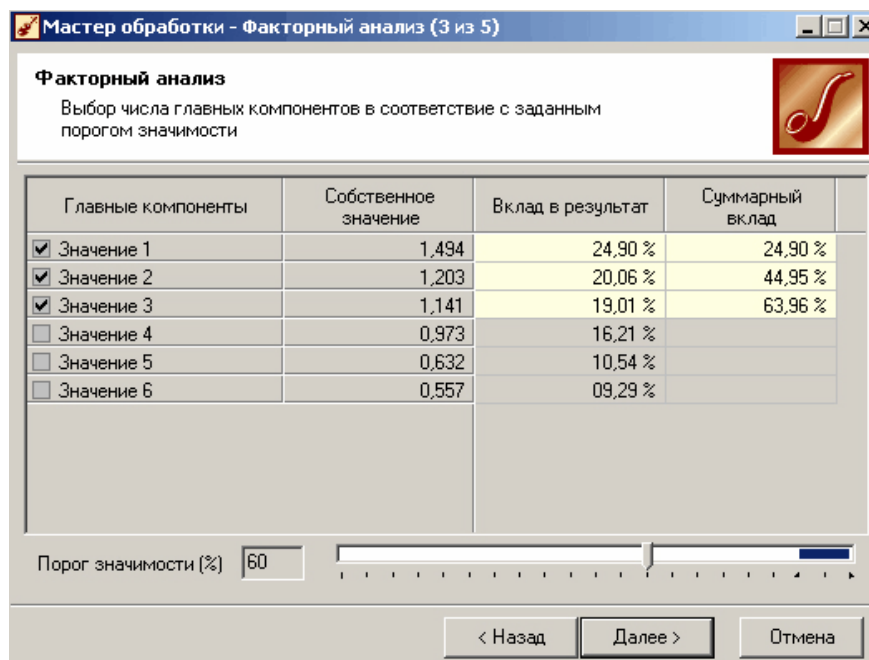
В Мастере обработки выберем факторный анализ и зададим входные поля

— "Личный доход в месяц", "Сумма кредита", "Стаж работы", "Рыночная стоимость автомобиля", "Рыночная стоимость недвижимости".



На следующем шаге предлагается запустить процесс понижения

размерности пространства входных факторов. После завершения процесса можно выбрать, какие из полученных в результате обработки факторы оставить для дальнейшей работы. Это делается путем указания необходимого порога значимости.



3. Результат

Теперь необходимо перейти на следующий шаг и выбрать способ визуализации; результаты просмотрим в таблице.

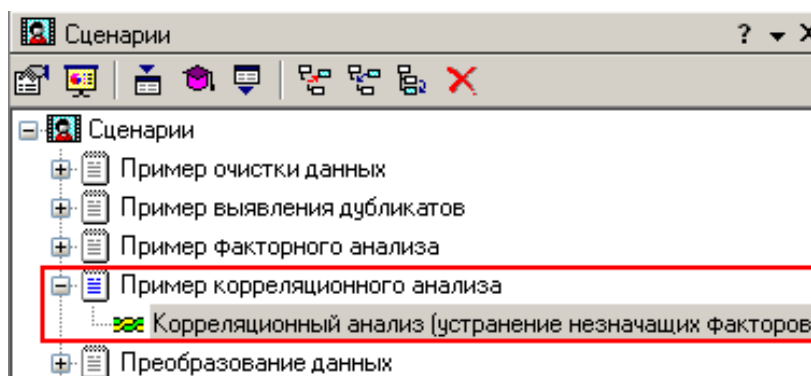
Фактор_1	Фактор_2	Фактор_3	Имя	Фамилия	Отчество
0,547	-1,472	-0,455	Николай	Абаджев	Васильевич
0,421	1,500	0,037	Светлана	Широкова	Николаевна
1,443	1,138	0,816	Вячеслав	Попков	Леонидович
-0,043	-0,184	-1,285	Юрий	Беляев	Алефтинович
0,906	0,595	-1,119	Аркадий	Репников	Ильич
-0,583	0,651	0,656	Анатолий	Калугин	Алексеевич
-1,115	-0,366	0,081	Нина	Смольникова	Дмитриевна
-0,650	-0,951	-0,677	Андрей	Катков	Викторович
0,392	0,943	0,004	Александр	Абаев	Викторович
-0,926	-0,470	1,168	Рустам	Кудабаев	Альбертович
-0,864	0,370	-0,320	Анна	Кондратьева	Васильевна
2,870	-1,104	1,053	Мила	Стрелкова	Николаевна
-1,217	0,359	-0,596	Алена	Копылова	Николаевна
-0,166	0,723	0,084	Николай	Кардапольцев	Георгиевич
2,266	2,454	2,623	Виталий	Шуклин	Георгиевич
-0,102	0,454	-0,571	Марина	Орлова	Анатолевна
-1,306	1,184	-0,224	Светлана	Меньшутина	Николаевна
-1,539	-0,718	0,649	Лариса	Корнилова	Владимировна
1,033	0,153	-0,997	Анатолий	Николаев	Николаевич
-0,104	-1,000	-1,173	Файруза	Миргалиева	Гимрановна
1,792	0,295	0,693	Николай	Иванов	Данилович

Практическая работа 4. Корреляционный анализ

Корреляционный анализ применяется для оценки зависимости выходных полей данных от входных факторов и устранения незначущих факторов. Принцип корреляционного анализа

состоит в поиске таких значений, которые в наименьшей степени коррелированы (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.

Покажем, как реализован данный фрагмент сценария в проекте "Демопример анализа данных.ded".

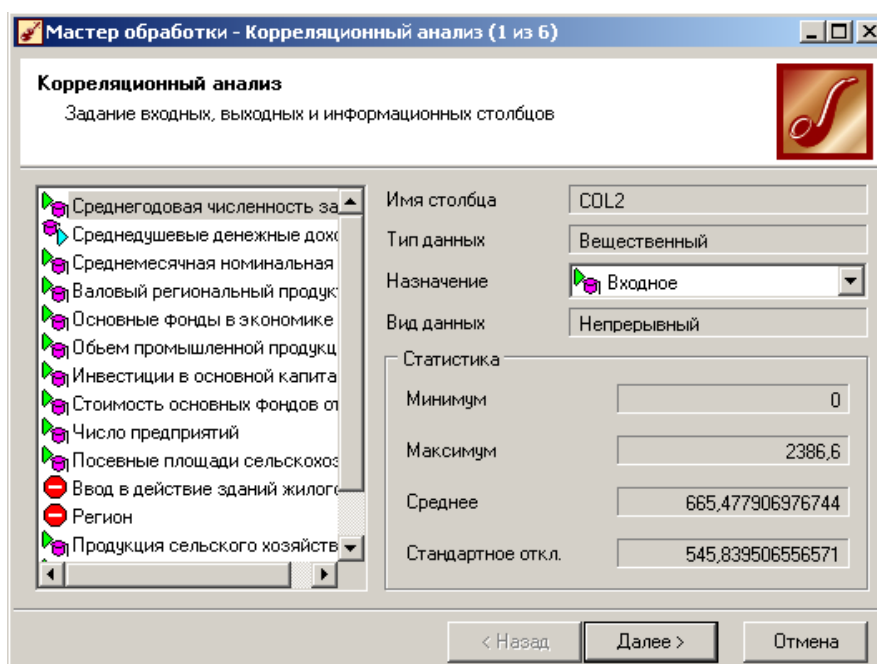


1. Исходные данные

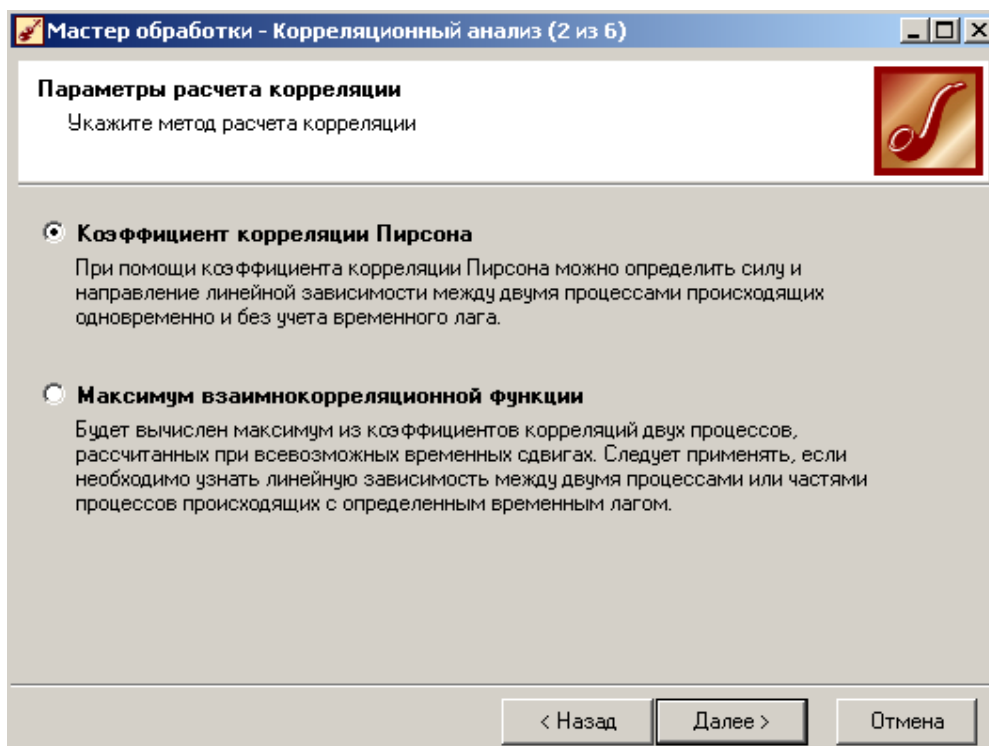
Рассмотрим применение обработки на примере данных из файла "region.txt". В данном примере определим степень влияния экономических показателей региона на среднедушевой доход жителей.

2. Устранение незначимых входных факторов

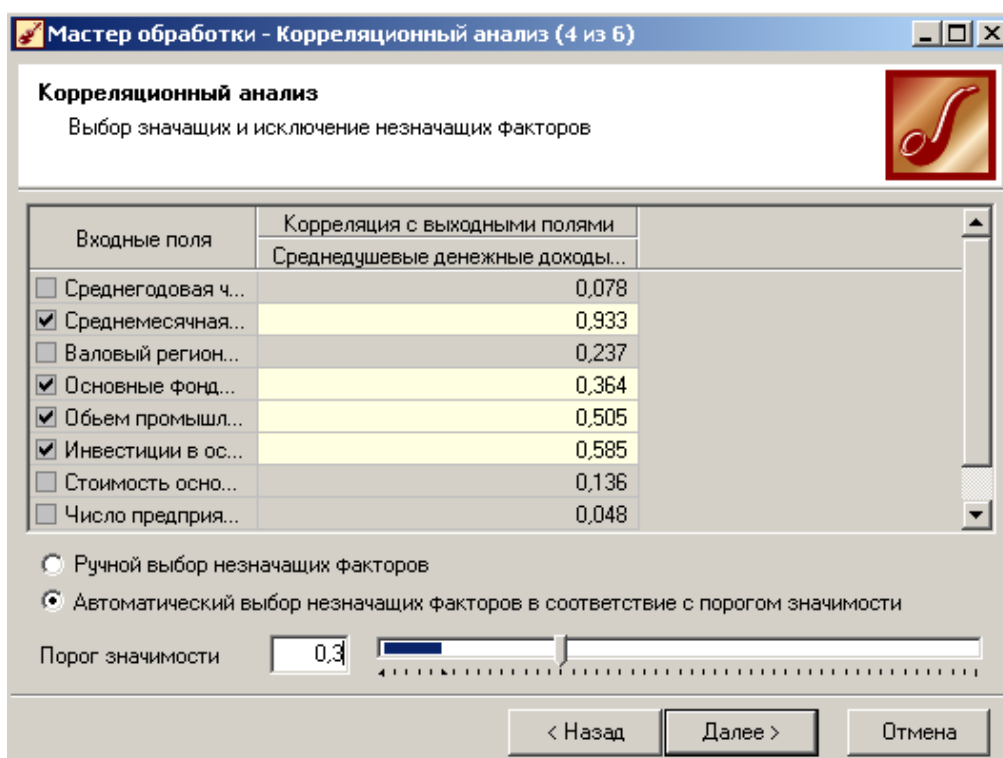
В Мастере обработки выбираем корреляционный анализ и задаем входные и выходные поля. Входными факторами будут являться все экономические показатели региона, а выходным будет "Среднедушевой денежный доход"



На следующем шаге необходимо выбрать метод на основе которого будет происходить расчет коэффициентов корреляции, выберем метод коэффициент корреляции Пирсона.



После выполнения предварительных настроек запускаем процесс корреляционного анализа, по результатам которого предлагается выбрать, какие факторы оставить для дальнейшей работы. Это делается либо вручную, основываясь на значениях матрицы ковариации, либо путем указания порога значимости (по умолчанию порог значимости равен 0.05).



3. Результат

По полученной матрице корреляции видно, какие факторы влияют сильнее, чем другие, и какие можно не учитывать при построении всевозможных моделей.

Матрица корреляции		
Входные поля		Корреляция с выходными полями
№	Поле	Среднедушевые денежные доход... ▲
2	Среднемесячная номинальная на...	0,933
12	Соотношение мужчин и женщин# (...)	-0,599
6	Инвестиции в основной капитал (...)	0,585
5	Объем промышленной продукции (...)	0,505
4	Основные фонды в экономике поп...	0,364
3	Валовый региональный продукт, (...)	0,237
11	Оборот розничной торговли (млн#...	0,235
9	Посевные площади сельскохозяй...	-0,175
7	Стоимость основных фондов отра...	0,136
10	Продукция сельского хозяйства# ...	-0,131
1	Среднегодовая численность занят...	0,078
8	Число предприятий	0,048

Практическая работа 5. Дубликаты и противоречия

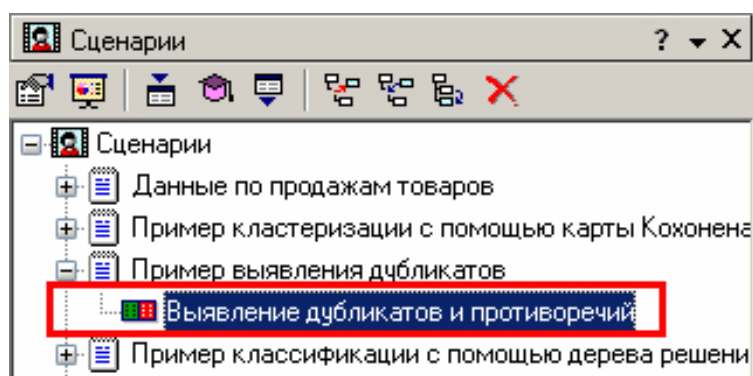
Одна из серьезных проблем, часто встречающаяся на практике, - наличие в данных дубликатов и противоречий.

Противоречивыми являются группы записей, в которых содержатся строки с одинаковыми входными факторами, но разными выходными. В такой ситуации непонятно, какое результирующее значение верное. Если противоречивые данные использовать для построения модели, то она окажется неадекватной. Поэтому противоречивые данные чаще всего лучше вообще исключить из исходной выборки.

Также в данных могут встречаться записи с одинаковыми входными факторами и одинаковыми выходными, т.е. дубликаты. Таким образом, данные несут избыточность. Присутствие дубликатов в анализируемых данных можно рассматривать как способ повышения "значимости" дублирующейся информации. Иногда они даже необходимы, например, если при построении модели нужно особо выделить некоторые наборы значений. Но все равно включение в выборку дублирующей информации должно происходить осознанно: в большинстве случаев дубликаты в данных являются следствием ошибок при подготовке данных. Так или иначе возникает задача выявления дубликатов и противоречий. В Deductor Studio для автоматизации этого процесса есть соответствующий инструмент – обработка "Дубликаты и противоречия".

Суть обработки состоит в том, что определяются входные (факторы) и выходные (результаты) поля. Алгоритм ищет во всем наборе записи, для которых одинаковым входным полям соответствуют одинаковые (дубликаты) или разные (противоречия) выходные поля. На основании этой информации создаются два дополнительных логических поля – "Дубликат" и "Противоречие", принимающие значения "правда" или "ложь". В дополнительные числовые поля "Группа дубликатов" и "Группа противоречий" записываются номер группы дубликатов и группы противоречий, в которые попадает данная запись. Если запись не является дубликатом или противоречием, то соответствующее поле будет пустым.

Покажем на примере, описанном в указанном фрагменте сценария проекта *"Демонстример анализа данных.ded"*.



1. Исходные данные

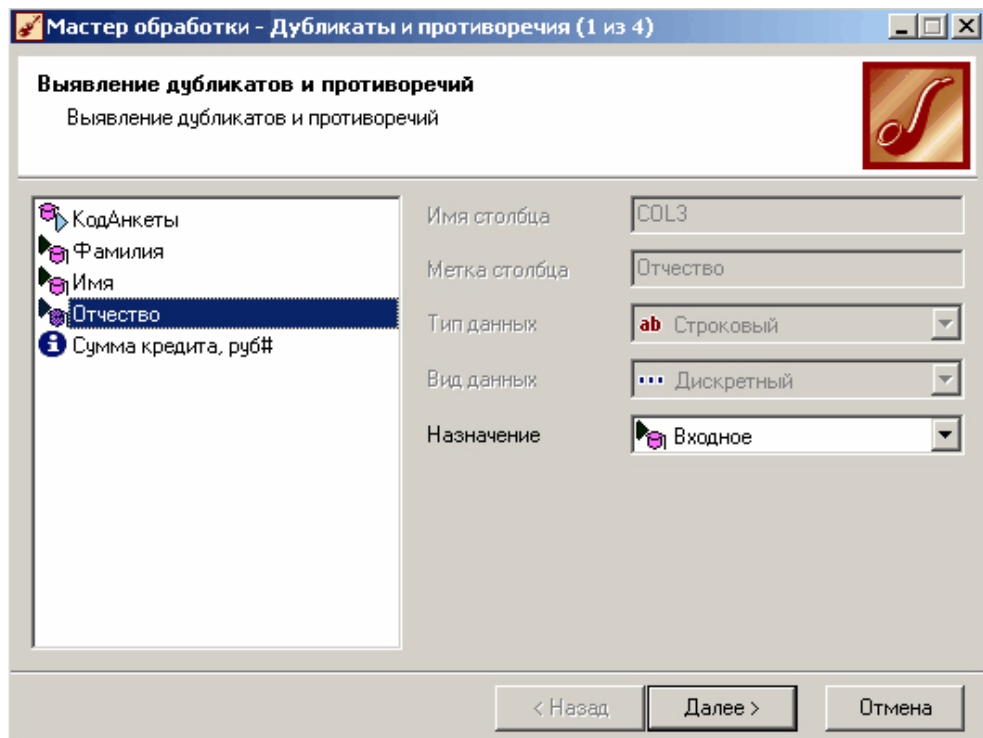
Рассмотрим механизм выявления дубликатов на примере данных файла "Anketa.txt". В этом файле находится информация об анкетных данных граждан, участвующих в кредитовании. Попробуем вычислить присутствие дубликатов. Импортируем данные из текстового файла и посмотрим их в виде таблицы.

Код Анкеты	Фамилия	Имя	Отчество	Сумма кредита, руб#
3072	Николаев	Анатолий	Николаевич	54000
3073	Миргалиева	Файруза	Гимрановна	42000
3074	Иванов	Николай	Данилович	51000
3076	Полякова	Тамара	Ивановна	36000
3999	Семакин	Николай	Анатолевич	60000
4000	Ларина	Елена	Валентиновна	53000
4000	Широкова	Светлана	Николаевна	54000
4001	Наговицын	Николай	Георгиевич	41000
4002	Бочкарев	Рудольф	Алексеевич	31000
4002	Симонова	Ольга	Сергеевна	70000
4004	Петров	Антон	Сергеевич	66000
4005	Чешкова	Елена	Борисовна	41000
4006	Видеева	Екатерина	Анатолевна	25000
4007	Рябова	Елена	Валентиновна	44000
4008	Касаткин	Сергей	Петрович	62000
4009	Масленников	Сергей	Алексеевич	65000
4010	Копылова	Алена	Николаевна	47000
4010	Терновой	Дмитрий	Сергеевич	60000
4012	Алексенко	Дмитрий	Дмитриевич	64000
4013	Ханнаков	Медахат	Рифкатович	120000
4015	Захаров	Алексей	Михайлович	37000
4016	Казанцев	Борис	Семенович	41000
4018	Перминова	Маргарита	Владимировна	26000

2. Поиск дубликатов и противоречий

Для выявления дубликатов запустим Мастер обработки. В нем выберем тип обработки "Дубликаты и противоречия".

На втором шаге Мастера необходимо настроить назначение полей.



На следующем шаге необходимо запустить процесс обработки.

3. Результат

После завершения выявления дубликатов посмотрим результат в виде таблицы дубликатов и противоречий.

В первом случае видно, что существуют одинаковые строки, являющиеся дубликатами. Данный обработчик показывает дубликаты и их принадлежность к группам дубликатов.

Дубликаты и противоречия						
1 / 8						
Дубликаты		Выходные поля	Входные поля			
Признак	Группа	Код Анкеты	Фамилия	Имя	Отчество	
<input checked="" type="checkbox"/>	2	3056	Калугин	Анатолий	Алексеевич	
<input checked="" type="checkbox"/>	2	3056	Калугин	Анатолий	Алексеевич	
<input checked="" type="checkbox"/>	3	3076	Полякова	Тамара	Ивановна	
<input checked="" type="checkbox"/>	3	3076	Полякова	Тамара	Ивановна	
<input checked="" type="checkbox"/>	4	4000	Широкова	Светлана	Николаевна	
<input checked="" type="checkbox"/>	4	4000	Широкова	Светлана	Николаевна	
<input checked="" type="checkbox"/>	1	4076	Бобров	Андрей	Владимирович	
<input checked="" type="checkbox"/>	1	4076	Бобров	Андрей	Владимирович	

Во втором случае видно, что при одинаковых "Фамилия", "Имя", "Отчество" оказываются различные Коды Анкет. В данном обработчике видно, у каких строк существуют противоречия и к какой группе они относятся.

Дубликаты и противоречия							
Дубликаты		Противоречия		Выходные поля	Входные поля		
	Признак	Признак	Группа	Код Анкеты	Фамилия	Имя	Отчество
<input type="checkbox"/>	<input checked="" type="checkbox"/>		1	3061	Абаев	Александр	Викторович
<input type="checkbox"/>	<input checked="" type="checkbox"/>		1	4026	Абаев	Александр	Викторович
<input type="checkbox"/>	<input checked="" type="checkbox"/>		2	4054	Евстафьев	Олег	Николаевич
<input type="checkbox"/>	<input checked="" type="checkbox"/>		2	4039	Евстафьев	Олег	Николаевич
<input type="checkbox"/>	<input checked="" type="checkbox"/>		3	4035	Ханнаков	Медахат	Рифкатович
<input type="checkbox"/>	<input checked="" type="checkbox"/>		3	4013	Ханнаков	Медахат	Рифкатович

Практическая работа 6. Трансформация данных

Часть примеров, входящих в группу "Трансформация данных", показывает, как с помощью инструментов Deductor Studio можно добиться тех или иных промежуточных задач, касающихся сбора аналитической информации, т. е. как выполнить трансформацию данных.

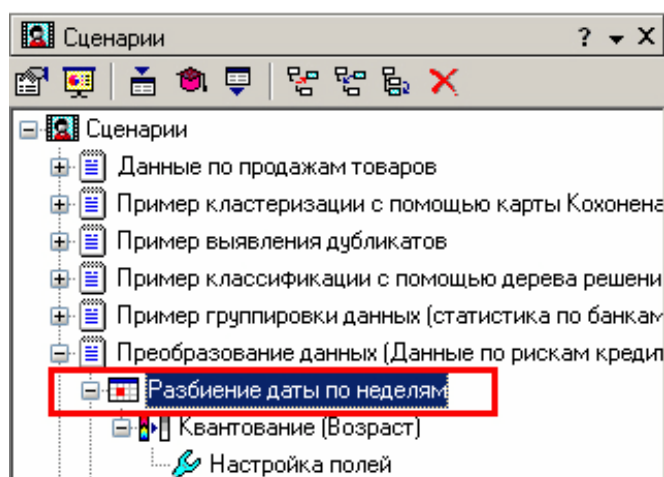
В приведенном демопримере рассматриваются такие задачи, как:

1. Разбиение даты
2. Квантование
3. Фильтрация данных
4. Группировка данных
5. Преобразование данных к скользящему окну
6. Настройка набора данных
7. Замена значений
8. Слияние

Разбиение даты

Разбиение даты служит для анализа всевозможных временных интервалов на основе имеющейся информации о дате и времени. Суть разбиения заключается в том, что на основе столбца с информацией о дате формируется другой столбец, в котором указывается, к какому заданному интервалу времени принадлежит строка данных. Тип интервала задается аналитиком, исходя из того, что он хочет получить: данные за год, квартал, месяц, неделю, день или сразу по всем интервалам.

Покажем, как выполнялась данная часть сценария проекта "Демопример анализа данных.ded".

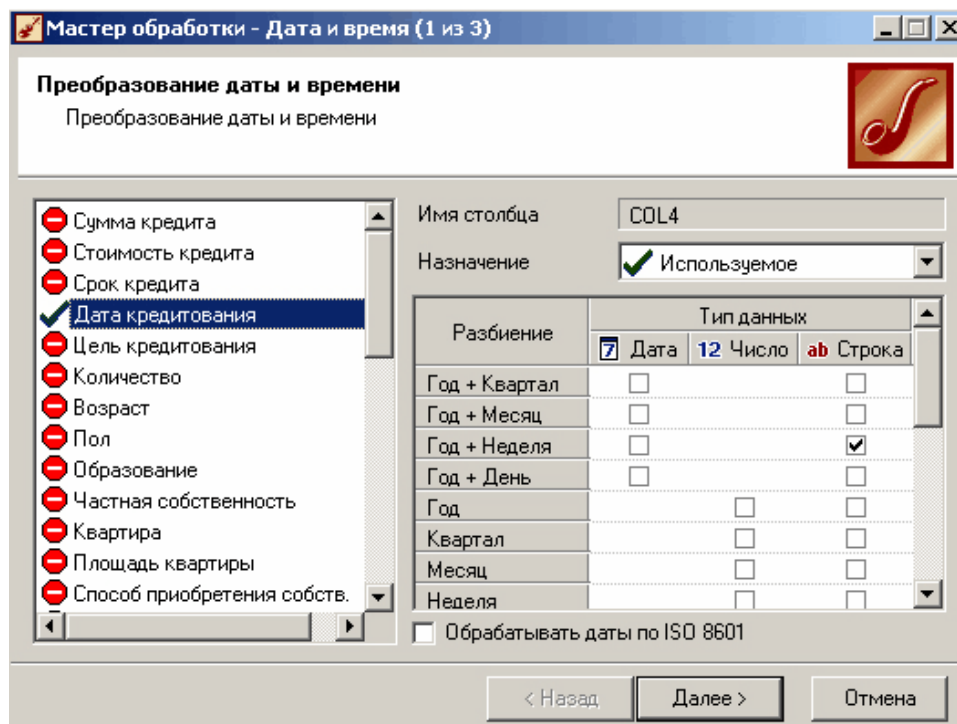


1. Исходные данные

Пусть нам необходимо получить данные по суммам взятых кредитов по неделям (в файле "Credit.txt" содержится информация за первые две недели 2003 года).

2. Обработка полученных данных

Для этого в Мастере обработки "Дата и Время" на втором шаге выберем поле "Дата кредитования" используемым, в появившейся после этого таблице настроек выберем назначение "Используемое" в столбце "Строка" напротив строки "Год + Неделя".



Больше никакие настройки не понадобятся, поэтому перейдем далее к выбору типа визуализации.

3. Результат

Выберем в качестве визуализатора "Куб".

В Мастере настройки полей куба выберем в качестве измерения появившийся после обработки столбец "Дата кредитования_YWStr (Год + Неделя)" и столбец "Цель кредитования", а в качестве факта – "Сумма кредита". Остальные поля сделаем неиспользуемыми.

Цель кредитования	2003-W01	2003-W02	Итого:
Иное	358 000,00	137 000,00	495 000,00
Оплата за образование	62 000,00	312 000,00	374 000,00
Оплата услуг (мед., юрид. и т.п.)	110 500,00	191 000,00	301 500,00
Покупка и ремонт недвижимости	404 000,00	538 000,00	942 000,00
Покупка товара	642 000,00	643 500,00	1 285 500,00
Турпоездки, развлечения и т.п.	35 500,00	113 500,00	149 000,00
Итого:	1 612 000,00	1 935 000,00	3 547 000,00

На следующем шаге перенесем одно измерение из области "доступных" в область "Измерения в строках", а другое – в область "Измерения в столбцах". Таким образом, на кросс-диаграмме имеем суммы взятых кредитов по неделям (за первые две недели года) в разрезе целей кредитования.

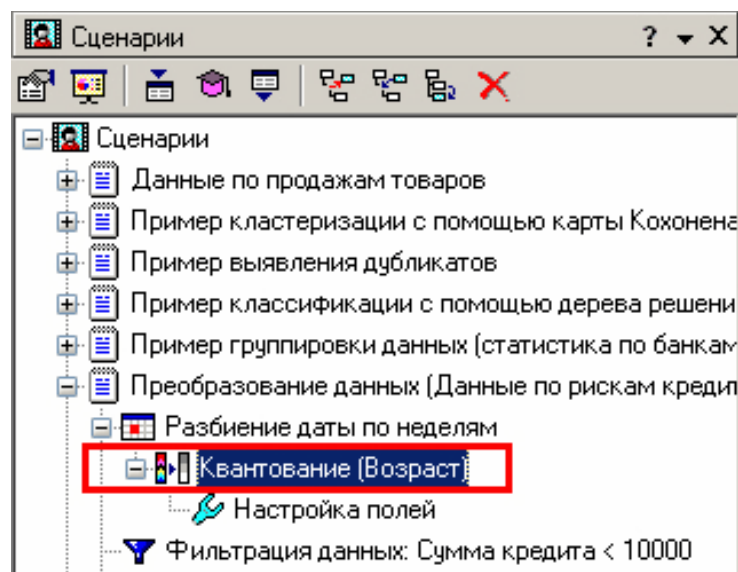
Квантование

Часто аналитику необходимо отнести непрерывные данные (например, Количество продаж) к какому-либо конечному набору (например, всю совокупность данных о количестве продаж необходимо разбить на 5 интервалов – от 0 до 100, от 100 до 200 и т.д. и отнести каждую запись исходного набора к какому-то конкретному интервалу) для анализа или фильтрации, исходя именно из этих интервалов. Для этого в Deductor Studio применяется инструмент квантования (или дискретизации).

Квантование предназначено для преобразования непрерывных данных в дискретные. Преобразование может проходить как по интервалам (данные разбиваются на заданное количество интервалов одинаковой длины), так и по квантилям (данные разбиваются на интервалы разной длины так, чтобы в каждом интервале находилось одинаковое количество записей). В качестве значений результирующего набора данных могут выступать номер интервала, нижняя или верхняя граница интервала, середина интервала либо метка интервала (значения, определяемые аналитиком).

Примером использования данного инструмента может служить разбиение данных о возрасте кредиторов на 5 интервалов (до 30 лет, от 30 до 40, от 40 до 50, от 50 до 60, старше 60 лет). Исходные данные распределятся по пяти интервалам именно так, поскольку согласно статистике, минимальное значение возраста кредитора 19, а максимальное - 69 лет. Это необходимо аналитику для оценки кредиторской активности разных возрастных групп с целью принятия решения о стимулировании кредиторов в группах с низкой активностью (например, уменьшение стоимости кредита для этих групп) и, быть может, увеличение прибыли в возрастных группах кредиторов с высоким риском (путем предложения дополнительных платных услуг). Причем аналитик желает видеть данные в разрезе по неделям.

Необходимая часть сценария находится в подсвеченной области проекта "Демопример анализа данных.ded".



1. Исходные данные

Воспользуемся данными, полученными при разбиении даты файла "Credit.txt".

2. Обработка данных

Воспользуемся Мастером квантования.

В нем выберем назначение поля "Возраст" используемым, укажем способ разбиения "По интервалам", зададим количество интервалов равное 5, в качестве значения выберем "Метка интервала".

На следующем шаге Мастера определим сами метки соответственно возраста кредиторов: "до 30 лет", "от 30 до 40 лет" и т.д.

Мастер обработки - Квантование (1 из 4)

Квантование
Настройка параметров квантования

Сумма кредита
Стоимость кредита
Срок кредита
Дата кредитования
Дата кредитования (Год + Неделя)
Цель кредитования
Количество
Возраст
Пол
Образование
Частная собственность
Квартира
Площадь квартиры

Имя столбца: COL7
Тип данных: Вещественный
Назначение: Используемое
Способ: По интервалам
Интервалов: 5
Значение: Метка интервала
Вид данных: Дискретный
Минимум: 19
Максимум: 70
Стандартное откл.: 12,9295541519765

< Назад Далее > Отмена

В нем выберем назначение поля "Возраст" используемым, укажем способ разбиения "По интервалам", зададим количество интервалов равное 5, в качестве значения выберем "Метка интервала".

На следующем шаге Мастера определим сами метки соответственно возраста кредиторов: "до 30 лет", "от 30 до 40 лет" и т.д.

Мастер обработки - Квантование (2 из 4)

Границы и метки интервалов
Настройка границ и меток интервалов квантования

Столбцы	
Имя	Интервалов
Возраст	5

Интервалы (изменены)		
N°	Граница	Метка
	0	
0	29	До 30 лет
1	39	От 30 до 40 лет
2	49	От 40 до 50 лет
3	59	От 50 до 60 лет
4	1000	Старше 60 лет

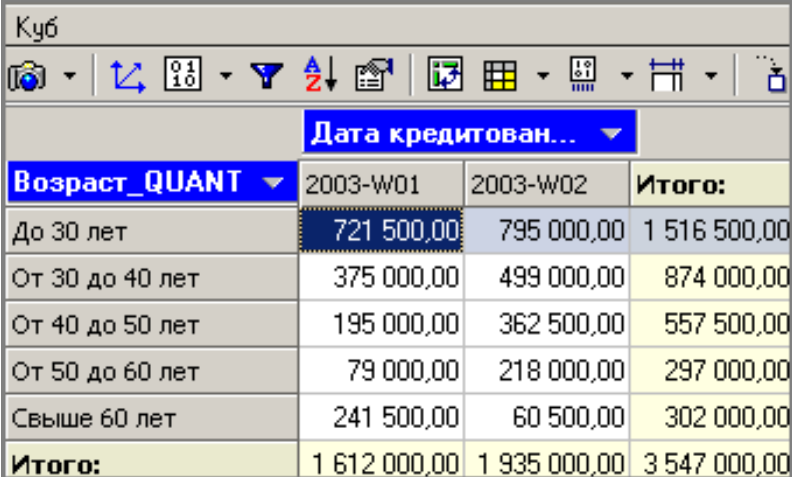
< Назад Далее > Отмена

3. Результат

После обработки выберем в качестве способа отображения "Куб". В Мастере укажем "Сумма кредита" в качестве факта, "Возраст" и поле "Дата кредитования (Год + Неделя)" в качестве измерения, остальные поля укажем неиспользуемыми.

Далее перенесем "Возраст" из доступных измерений в "Измерения в строках", а "Дата кредитования (Год + Неделя)" в "Измерения в столбцах".

На кросс-диаграмме теперь видна информация о том, какие суммы кредитов берут кредиторы определенных возрастных групп по неделям.



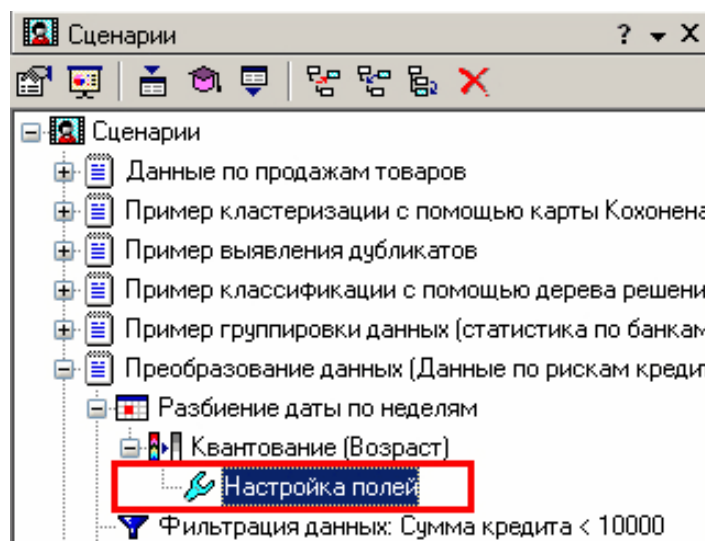
	Дата кредитован...		
Возраст_QUANT	2003-W01	2003-W02	Итого:
До 30 лет	721 500,00	795 000,00	1 516 500,00
От 30 до 40 лет	375 000,00	499 000,00	874 000,00
От 40 до 50 лет	195 000,00	362 500,00	557 500,00
От 50 до 60 лет	79 000,00	218 000,00	297 000,00
Свыше 60 лет	241 500,00	60 500,00	302 000,00
Итого:	1 612 000,00	1 935 000,00	3 547 000,00

Настройка набора данных

Настройка набора данных применяется, когда необходимо изменить имя, метку, размер, тип, вид и назначение полей текущей таблицы данных для более удобного дальнейшего использования. Кроме того, в данном обработчике реализован механизм кэширования данных, который позволяет загрузить данные в оперативную память и оптимизировать скорость выполнения сценариев.

Замечание: Данный обработчик аналогичен шагу Мастера настройки полей при импорте данных в программу, рассмотренному выше.

Покажем, как выполняется данный обработчик в сценарии, находящийся в проекте "Демонстрация анализа данных.ded".

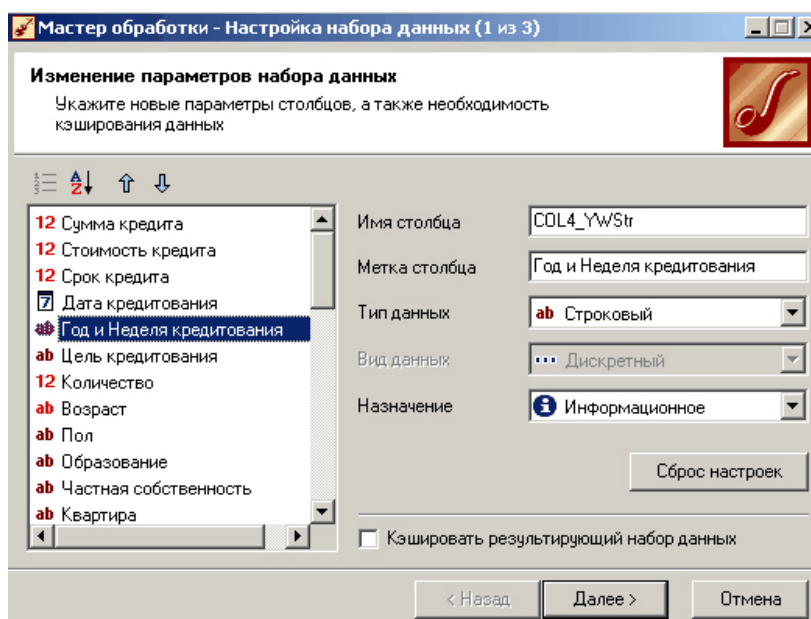


1. Исходные данные

Продemonстрируем использование настройки полей, используя данные, полученные после квантования возраста кредиторов на интервалы из примера выше. Пусть необходимо изменить метку поля "Дата кредитования (Год + Неделя)" на более информативную при подготовке отчетности - "Год и неделя кредитования". Пусть также для дальнейшего использования необходимо установить размер поля "Цель кредитования" 30 символов и использовать поле "Срок кредита" как дискретное.

2. Выполнение настройки

В Мастере настройки выделим столбец "Дата кредитования (Год +Неделя)" и укажем ему новую метку. Подобные действия по изменению произведем и с другими полями.



3. Результат

После настройки полей полученный отчет, представленный в виде кросс- таблицы, будет выглядеть следующим образом:

		Давать кредит		
		Да	Нет	Итого:
2003-W01	6	194 500,00	7 500,00	202 000,00
	12	244 500,00	176 500,00	421 000,00
	18	48 000,00	249 000,00	297 000,00
	24		195 500,00	195 500,00
	30		151 000,00	151 000,00
	36		229 000,00	229 000,00
	42		58 500,00	58 500,00
	48		58 000,00	58 000,00
	Итого:		487 000,00	1 125 000,00
2003-W02	6	183 500,00	25 500,00	209 000,00
	12	109 000,00	373 500,00	482 500,00
	18	61 000,00	440 000,00	501 000,00
	24	37 000,00	319 000,00	356 000,00
	30		176 500,00	176 500,00
	36		156 500,00	156 500,00
	42		53 500,00	53 500,00
Итого:		390 500,00	1 544 500,00	1 935 000,00
Итого:		877 500,00	2 669 500,00	3 547 000,00

Как видно, требуемый результат был достигнут.

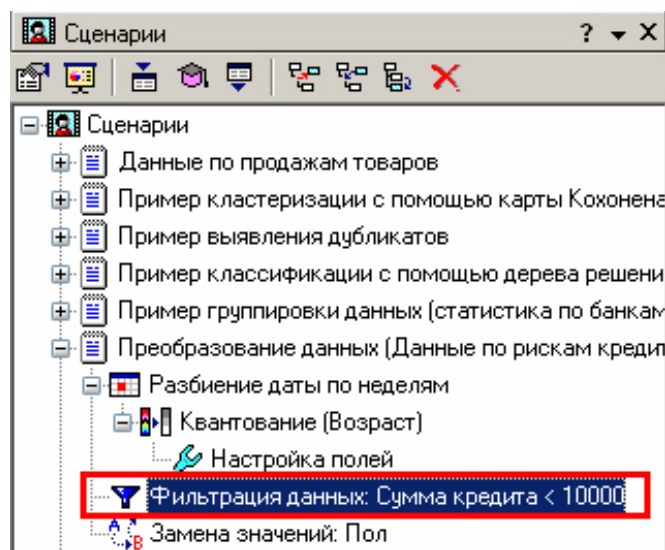
Фильтрация данных

Почти всегда исходный набор данных или набор данных после обработки аналитику нужно отфильтровать. Фильтрация бывает необходима для разбиения данных на какие-либо группы (например, товарные) для последующей обработки или анализа данных уже отдельно по каждой группе. Также некоторые данные могут не подходить для дальнейшего анализа в силу накладываемых условий (например, если на каком-либо этапе обработки данных были выявлены противоречивые записи, то их следует исключить из последующей обработки). В этом случае также возникает необходимость фильтрации записей.

Фильтрация позволяет из исходного набора данных получить новую таблицу, удовлетворяющую условиям, которые определены аналитиком. В Deductor Studio механизм построения условий фильтрации прост для понимания. В окне Мастера можно задать несколько элементарных условий фильтрации (<ПОЛЕ> <ОТНОШЕНИЕ> <ЗНАЧЕНИЕ>), последовательно связанных логическими операциями (И, ИЛИ).

Рассмотрим ситуацию, когда аналитику необходимо оценить кредитоспособность потенциального заемщика. Предполагается, что заемщики, берущие суммы разного диапазона, ведут себя по-разному, следовательно, модели прогноза должны быть свои для каждой группы, т. е. для дальнейшего построения моделей оценки кредитоспособности различных категорий заемщиков необходимо использовать фильтрацию.

Возможность фильтрации демонстрируется в выделенной части проекта "Демонстример анализа данных.ded".



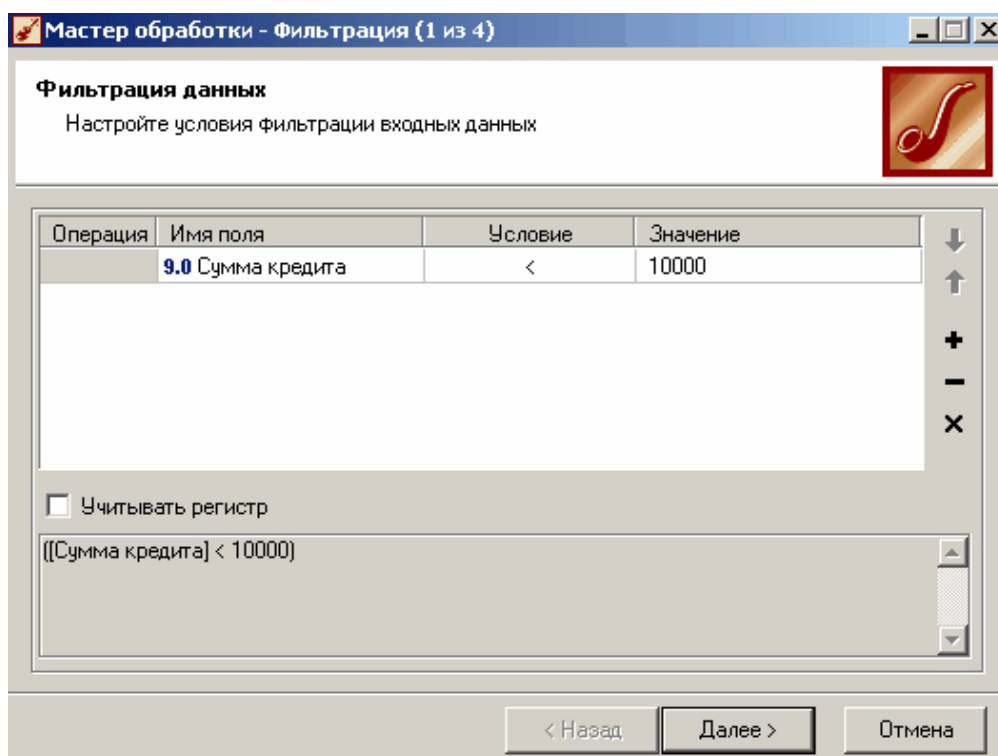
1. Исходные данные

Воспользуемся данными файла "Credit.txt".

2. Выполнение фильтрации данных

Определим для примера группу заемщиков, взявших кредит менее 10000 руб.

Для этого, находясь на узле импорта данных, из текстового файла запустим Мастер обработки. В нем в качестве метода обработки выберем фильтрацию. На втором шаге Мастера задаем условия фильтрации (при необходимости их можно добавлять или удалять соответствующими кнопками на форме). Поскольку необходимо отфильтровать данные только по заемщикам, взявшим кредит менее 10000, то в графе "Имя поля" выбираем поле "Сумма кредита", в графе "Условие" - знак меньше, а в графе "Значение" пишем "10000".



Больше никаких условий не требуется, поэтому переходим на следующий шаг Мастера и запускаем процесс фильтрации. После выполнения обработки можно манипулировать уже только с данными по заемщикам выбранного кредитного диапазона.

3. Результат

В правильности выполненной операции можно легко убедиться, выбрав в качестве визуализации данных статистику и просмотрев значения минимального и максимального значения поля "Сумма кредита".

	Метка столбца	Статистика: Кол-во значен...				
		Мини...	Макс...	Среднее	Стан...	Σ C
1	9.0 Сумма кредита	2000	9500	5933,333333333333	0503352841	1
2	9.0 Стоимость кред...	400	1900	1186,666666666667	1100670568	
3	9.0 Срок кредита	6	6	6	0	
4	7 Дата кредитова...	01.01.2003	11.01.2003	05.01.2003 19:12:00	здн. 9:35:59	
5	ab Цель кредитова...	4	33	19,7	10,927	
6	9.0 Количество	1	1	1	0	
7	9.0 Возраст	20	67	36,23333333333333	1340531625	
8	ab Пол	3	3	3	0	
9	ab Образование	6	11	9,367	2,168	
10	ab Частная собств...	2	3	2,267	0,442	
11	ab Кредит...	2	2	2,4	0,48	

Критерии оценивания устных ответов аспирантов

Развернутый ответ аспиранта должен представлять собой связное, логически последовательное сообщение на определенную тему, показывать его умение применять определения, правила в конкретных случаях.

При оценке устного ответа аспиранта необходимо руководствоваться следующими критериями:

- 1) полнота и правильность ответа;
- 2) степень осознанности, понимания изучаемого материала;
- 3) знание терминологии и правильное ее использование;
- 4) соответствие требованиям рабочей программы по дисциплине.

Оценка «зачтено» за устный ответ ставится, если аспирант:

- 1) ориентируется в излагаемом материале, владеет базовой терминологией в объеме, предусмотренном рабочей программой дисциплины;
- 2) обнаруживает понимание материала, может обосновать свои суждения, подкрепляет теоретические положения примерами;
- 3) умеет структурировать содержание ответа в соответствии с поставленным вопросом;
- 4) не допускает (или допускает немногочисленные негрубые) ошибки при анализе языковых фактов; способен исправить допущенные им ошибки при помощи уточняющих вопросов преподавателя.

Порядок проведения дифференцированного зачета

Дифференцированный зачет используется для оценки соответствия результатов освоения дисциплины аспирантом планируемыми.

Дифференцированный зачет проводится путем оценивания представления аспирантом индивидуального задания.

Задание выдается преподавателем и состоит из письменного выполнения следующих элементов:

- индивидуальный план работы преподавателя (фрагмент за семестр по одной дисциплине);
- календарный план занятий по дисциплине на семестр;
- рабочая программа дисциплины (фрагмент);
- план проведения занятия (любой формы);
- презентация занятия.

Аспирант в установленный преподавателем срок сдает преподавателю выполненное индивидуальное задание для проверки. При положительном результате проверки аспирант представляет презентацию и обсуждает выполненное индивидуальное задание с преподавателем, по итогам презентации и обсуждения преподаватель выставляет оценку. Оценка объявляется аспиранту и заносится в зачетную ведомость.

Выполненные индивидуальные задания в электронном виде и на бумажном носителе хранятся на кафедре системного анализа и управления.

Критерии и процедура оценивания результатов дифференцированного зачета

Оценки за представление аспирантом индивидуального задания выставляются, исходя из следующих критериев:

- «**отлично**»: если аспирант глубоко и прочно усвоил весь программный материал лекций и демонстрирует это в задании, все документы выполнены без ошибок, последовательно, грамотно и логически построены, излагает свои решения, хорошо их объясняя и обосновывая;
- «**хорошо**»: если аспирант твердо знает программный материал, не допускает существенных неточностей в его изложении, использует ограниченный круг источников, вместо своего решения в задании излагает одно из стандартных;
- «**удовлетворительно**»: если аспирант поверхностно усвоил основной материал лекций, не знает деталей, допускает неточности, при разработке задания привлекает мало оригинального материала, пользуясь, в основном, стандартными решениями и формулировками;
- «**неудовлетворительно**»: если аспирант не знает значительной части программного материала, в задании допущены существенные ошибки, с большими затруднениями выполняет или, по существу, не выполняет задания, не может его объяснить.

ПЕРЕЧЕНЬ УЧЕБНОЙ ЛИТЕРАТУРЫ И РЕСУРСОВ СЕТИ «ИНТЕРНЕТ»

. Основная литература

1. Ясницкий, Л.Н. Интеллектуальные системы [Электронный ресурс]: учебник / Л.Н. Ясницкий. – эл. изд. – М.: Лаборатория знаний, 2016. – 224 с.
(http://biblioclub.ru/index.php?page=book_red&id=119452)
2. Статистические методы анализа данных [Электронный ресурс]: учебник / Л.И. Ниворожкина, С.В. Арженовский, А.А. Рудяга [и др.]; под общ. ред. д-ра экон. наук, проф. Л.И. Ниворожкиной. – М.: РИОР: ИНФРА-М, 2016. – 333 с.
(<http://znanium.com/catalog.php?bookinfo=556760>)
3. Информационные аналитические системы [Электронный ресурс]: учебник / Т.В. Алексеева, Ю.В. Амириди, В.В. Дик и др.; под ред. В.В. Дика. – М.: МФПУ Синергия, 2013. – 384 с. (<http://znanium.com/catalog.php?bookinfo=451186>)

Дополнительная литература

1. Романов, А.Н. Советующие информационные системы в экономике [Электронный ресурс]: учеб. пособие / А.Н. Романов, Б.Е. Одинцов. – М.: ИНФРА-М, 2017. – 485 с.
(<http://znanium.com/catalog.php?bookinfo=854392>)
2. Вейнберг Р.Р. Интеллектуальный анализ данных и систем управления бизнес-правилами в телекоммуникациях [Электронный ресурс]: монография / Р.Р. Вейнберг. – М.: НИЦ ИНФРА-М, 2016. – 173 с. (<http://znanium.com/catalog.php?bookinfo=520998>)
3. Нестеров, С.А. Интеллектуальный анализ данных средствами MS SQL Server 2008 [Электронный ресурс]: учебное пособие / С.А. Нестеров. – 2-е изд., испр. – М.: Национальный Открытый Университет «ИНТУИТ», 2016. – 338 с.
(http://biblioclub.ru/index.php?page=book_red&id=429083&sr=1)
4. Жуковский, О.И. Информационные технологии и анализ данных [Электронный ресурс]: учебное пособие / О.И. Жуковский; Министерство образования и науки Российской Федерации, Томский Государственный Университет Систем Управления и Радиоэлектроники (ТУСУР). – Томск: Эль Контент, 2014. – 130 с.
(http://biblioclub.ru/index.php?page=book_red&id=480500&sr=1)
5. Карпузова, В.И. Информационные технологии в менеджменте [Электронный ресурс]: учебное пособие / В.И. Карпузова, Э.Н. Скрипченко, К.В. Чернышева, Н.В. Карпузова. – 2-е изд., доп. – М.: Вузовский учебник: НИЦ ИНФРА-М, 2014. – 301 с.
(<http://znanium.com/catalog.php?bookinfo=410374>)
6. Чубукова, И.А. Data Mining [Электронный ресурс]: учебное пособие / И.А. Чубукова. – 2-е изд., испр. – М.: Интернет-Университет Информационных Технологий, 2008. – 383 с.
(http://biblioclub.ru/index.php?page=book_red&id=233055)

Учебно-методическое обеспечение самостоятельной работы аспиранта

- Методические указания для самостоятельной работы аспирантов;
- Методические указания по практическим занятиям.

Ресурсы сети «Интернет»

1. Информационная справочная система «Консультант плюс».
2. Библиотека ГОСТов www.gostrf.com.
3. Сайт Российской государственной библиотеки. <http://www.rsl.ru/>
4. Сайт Государственной публичной научно-технической библиотеки России. <http://www.gpntb.ru/>
5. Каталог образовательных интернет ресурсов <http://www.edu.ru/modules.php>
6. Электронные библиотеки: <http://www.pravoteka.ru/>, <http://www.zodchii.ws/>, <http://www.tehlit.ru/>
7. Специализированный портал по информационно-коммуникационным технологиям в образовании <http://www.ict.edu.ru>

Электронно-библиотечные системы:

- ЭБС издательства «Лань» <https://e.lanbook.com/>
- ЭБС издательства «Юрайт» <https://biblio-online.ru/>
- ЭБС «Университетская библиотека онлайн» <https://biblioclub.ru/>
- ЭБС «ZNANIUM.COM» <https://znanium.com>
 - ЭБС «IPRbooks» <https://iprbookshop.ru>
 - ЭБС «Elibrary» <https://elibrary.ru>
- Автоматизированная информационно-библиотечная система «Mark -SQL»
<https://informsystema.ru>
- Система автоматизации библиотек «ИРБИС 64» <https://elnit.org>

Информационные справочные системы:

1. Система ГАРАНТ: информационный правовой портал [Электронный ресурс]. – Электр.дан. <http://www.garant.ru/>
2. Консультант Плюс: справочно - поисковая система [Электронный ресурс]. – Электр.дан. www.consultant.ru/
3. ООО «Современные медиа-технологии в образовании и культуре». <http://www.informio.ru/>
4. Программное обеспечение Норма CS «Горное дело и полезные ископаемые» <https://softmap.ru/normacs/normacs-gornoe-delo-i-poleznye-iskopaemye/>
5. Информационно-справочная система «Техэксперт: Базовые нормативные документы» <http://www.cntd.ru/>