

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

## **МЕТОДИЧЕСКИЕ УКАЗАНИЯ**

для обучающихся по организации и проведению самостоятельной работы  
по дисциплине «Информационно-аналитические системы и технологии»  
направление подготовки 38.03.05 Бизнес-информатика  
направленность (профиль) «Информационная бизнес-аналитика и цифровые  
технологии»

## **ВВЕДЕНИЕ**

Самостоятельная работа - планируемая учебная, учебно-исследовательская, научно-исследовательская работа студентов, выполняемая во внеаудиторное (аудиторное) время по заданию и при методическом руководстве преподавателя, но без его непосредственного участия (при частичном непосредственном участии преподавателя, оставляющем ведущую роль за работой студентов).

Самостоятельная работа студентов в ВУЗе является важным видом учебной и научной деятельности студента. Самостоятельная работа студентов играет значительную роль в рейтинговой технологии обучения.

Целью данной дисциплины является формирование набора (ПК-4) компетенций будущего бакалавра по направлению подготовки 38.03.05 «Бизнес-информатика».

Задачами дисциплины «Информационно-аналитические системы и технологии» являются: изучение стратегических перспектив, увязывании финансовых и стратегических целей компании, разработке и формировании приоритета стратегических мероприятий, управлении реализацией стратегии компании, а также сокращении издержек и увеличении производительности.

## **Общая характеристика самостоятельной работы студента при изучении дисциплины «Информационно-аналитические системы и технологии»**

Целью самостоятельной работы студентов является овладение фундаментальными знаниями, профессиональными умениями и навыками деятельности по профилю, опытом творческой, исследовательской деятельности. Самостоятельная работа студентов способствует развитию самостоятельности, ответственности и организованности, творческого подхода к решению проблем учебного и профессионального уровня.

Задачами СР являются:

- систематизация и закрепление полученных теоретических знаний и практических умений студентов;
- углубление и расширение теоретических знаний;
- формирование умений использовать нормативную, правовую, справочную документацию и специальную литературу;
- развитие познавательных способностей и активности студентов: творческой инициативы, самостоятельности, ответственности и организованности;
- формирование самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;
- развитие исследовательских умений;
- использование материала, собранного и полученного в ходе самостоятельных занятий на семинарах, на практических и лабораторных занятиях, при написании курсовых и выпускной квалификационной работ.

### **План-график выполнения самостоятельной работы**

| <b>№</b> | <b>Наименование разделов и тем дисциплины, их краткое содержание; вид самостоятельной работы</b> | <b>Форма контроля</b>         |
|----------|--------------------------------------------------------------------------------------------------|-------------------------------|
|          | <b>8 семестр</b>                                                                                 |                               |
| 1        | Подготовка к лабораторной работе                                                                 | Выполнение лабораторных работ |
| 2        | Подготовка к лекции                                                                              | Опрос                         |
| 3        | Самостоятельное изучение литературы                                                              | Опрос                         |
| 4        | Выполнение курсовой работы                                                                       | Защита курсовой работы        |
| 5        | Подготовка к экзамену                                                                            | Сдача экзамена                |

### **Методические указания по изучению теоретического материала**

#### **Работа с книгой**

При работе с книгой необходимо подобрать литературу, научиться правильно ее читать, вести записи. Для подбора литературы в библиотеке используются алфавитный и систематический каталоги.

Важно помнить, что рациональные навыки работы с книгой - это всегда большая экономия времени и сил.

Правильный подбор учебников рекомендуется преподавателем, читающим лекционный курс. Необходимая литература может быть также указана в методических разработках по данному курсу.

Изучая материал по учебнику, следует переходить к следующему вопросу только после правильного уяснения предыдущего, описывая на бумаге все выкладки и вычисления (в том числе те, которые в учебнике опущены или на лекции даны для самостоятельного вывода).

При изучении любой дисциплины большую и важную роль играет самостоятельная индивидуальная работа.

Особое внимание следует обратить на определение основных понятий курса. Студент должен подробно разбирать примеры, которые поясняют такие определения, и уметь строить аналогичные примеры самостоятельно. Нужно добиваться точного представления о том, что изучаешь. Полезно составлять опорные конспекты. При изучении материала по учебнику полезно в тетради (на специально отведенных полях) дополнять конспект лекций. Там же следует отмечать вопросы, выделенные студентом для консультации с преподавателем.

Выводы, полученные в результате изучения, рекомендуется в конспекте выделять, чтобы они при перечитывании записей лучше запоминались.

Опыт показывает, что многим студентам помогает составление листа опорных сигналов, содержащего важнейшие и наиболее часто употребляемые формулы и понятия. Такой лист помогает запомнить формулы, основные положения лекции, а также может служить постоянным справочником для студента.

Различают два вида чтения; первичное и вторичное. *Первичное* - это внимательное, неторопливое чтение, при котором можно остановиться на трудных местах. После него не должно остаться ни одного непонятого слова. Содержание не всегда может быть понятно после первичного чтения.

Задача *вторичного* чтения полное усвоение смысла целого (по счету это чтение может быть и не вторым, а третьим или четвертым).

### **Правила самостоятельной работы с литературой**

Как уже отмечалось, самостоятельная работа с учебниками и книгами (а также самостоятельное теоретическое исследование проблем, обозначенных преподавателем на лекциях) - это важнейшее условие формирования у себя научного способа познания. Основные советы здесь можно свести к следующим:

- Составить перечень книг, с которыми Вам следует познакомиться;
- Сам такой перечень должен быть систематизированным.
- Обязательно выписывать все выходные данные по каждой книге (при написании курсовых и дипломных работ это позволит очень сэкономить время).
- Разобраться для себя, какие книги (или какие главы книг) следует прочитать более внимательно, а какие - просто просмотреть.
- При составлении перечней литературы следует посоветоваться с преподавателями и научными руководителями (или даже с более подготовленными и эрудированными сокурсниками), которые помогут Вам лучше сориентироваться, на что стоит обратить большее внимание, а на что вообще не стоит тратить время...
- Естественно, все прочитанные книги, учебники и статьи следует конспектировать, но это не означает, что надо конспектировать «все подряд»: можно выписывать кратко основные идеи автора и иногда приводить наиболее яркие и показательные цитаты (с указанием страниц).

Чтение научного текста является частью познавательной деятельности. Ее цель - извлечение из текста необходимой информации. От того на сколько осознанна читающим собственная внутренняя установка при обращении к печатному слову (найти нужные сведения, усвоить информацию полностью или частично, критически проанализировать материал и т.п.) во многом зависит эффективность осуществляемого действия.

Выделяют **четыре основные установки в чтении научного текста:**

1. информационно-поисковый (задача - найти, выделить искомую информацию);
2. усваивающая (усилия читателя направлены на то, чтобы как можно полнее осознать и запомнить как самисведения, излагаемые автором, так и всю логику его рассуждений);

3. аналитико-критическая (читатель стремится критически осмыслить материал, проанализировав его, определив свое отношение к нему);

4. творческая (создает у читателя готовность в том или ином виде - как отправной пункт для своих рассуждений, как образ для действия по аналогии и т.п. - использовать суждения автора, ход его мыслей, результат наблюдения, разработанную методику, дополнить их, подвергнуть новой проверке).

#### ***Основные виды систематизированной записи прочитанного:***

1. Аннотирование - предельно краткое связное описание просмотренной или прочитанной книги (статьи), ее содержания, источников, характера и назначения;

2. Планирование - краткая логическая организация текста, раскрывающая содержание и структуру изучаемого материала;

3. Тезирование - лаконичное воспроизведение основных утверждений автора без привлечения фактического материала;

4. Цитирование - дословное выписывание из текста выдержек, извлечений, наиболее существенно отражающих ту или иную мысль автора;

5. Конспектирование - краткое и последовательное изложение содержания прочитанного.

Конспект - сложный способ изложения содержания книги или статьи в логической последовательности. Конспект аккумулирует в себе предыдущие виды записи, позволяет всесторонне охватить содержание книги, статьи. Поэтому умение составлять план, тезисы, делать выписки и другие записи определяет и технологию составления конспекта.

#### **Методические указания по составлению конспекта**

1. Внимательно прочитайте текст. Уточните в справочной литературе непонятные слова. При записи не забудьте вынести справочные данные на поля конспекта.

2. Выделите главное, составьте план.

3. Кратко сформулируйте основные положения текста, отметьте аргументацию автора.

4. Законспектируйте материал, четко следуя пунктам плана. При конспектировании старайтесь выразить мысль своими словами. Записи следует вести четко, ясно.

5. Грамотно записывайте цитаты. Цитируя, учитывайте лаконичность, значимость мысли.

В тексте конспекта желательно приводить не только тезисные положения, но и их доказательства. При оформлении конспекта необходимо стремиться к емкости каждого предложения. Мысли автора книги следует излагать кратко, заботясь о стиле и выразительности написанного. Число дополнительных элементов конспекта должно быть логически обоснованным, записи должны распределяться в определенной последовательности, отвечающей логической структуре произведения. Для уточнения и дополнения необходимо оставлять поля.

Овладение навыками конспектирования требует от студента целеустремленности, повседневной самостоятельной работы.

#### **Методические указания по подготовке к экзамену**

В период подготовки к экзамену студенты вновь обращаются к пройденному учебному материалу. При этом они не только скрепляют полученные знания, но и получают новые. Подготовка студента к экзамену включает в себя три этапа:

-самостоятельная работа в течение семестра;

-непосредственная подготовка в дни, предшествующие экзамену по темам курса;

-подготовка к ответу на вопросы, содержащиеся в билетах.

Литература для подготовки к экзамену рекомендуется преподавателем либо указана в учебно-методическом комплексе. Для полноты учебной информации и ее сравнения лучше использовать не менее двух учебников. Студент вправе сам придерживаться любой из представленных в учебниках точек зрения по спорной проблеме (в том числе отличной от преподавателя), но при условии достаточной научной аргументации.

Основным источником подготовки к экзамену является конспект лекций, где учебный материал дается в систематизированном виде, основные положения его детализируются, подкрепляются современными фактами и информацией, которые в силу новизны не вошли в опубликованные печатные источники. В ходе подготовки к экзамену студентам необходимо обращать внимание не только на уровень запоминания, но и на степень понимания излагаемых проблем.

Экзамен проводится по билетам, охватывающим весь пройденный материал. По окончании ответа экзаменатор может задать студенту дополнительные и уточняющие вопросы. На подготовку к ответу по вопросам билета студенту дается 30 минут с момента получения им билета. Положительным также будет стремление студента изложить различные точки зрения на рассматриваемую проблему, выразить свое отношение к ней, применить теоретические знания по современным проблемам жилищного права.

Результаты экзамена объявляются студенту после окончания ответа в день сдачи.

## Задания для подготовки к лабораторным работам

### Задание 1. Создание ХД в DeductorStudio 5.3

- 1. Для создания нового хранилища данных или подключения к существующему нужно перейти на закладку **Подключения** и запустить **Мастер подключений**.
- 2. Пройти первые два шага, выбрав тип приемника (источника) DeductorWarehouse и тип базы данных Firebird.
- 3. На третьем шаге нужно задать параметры базы данных, в которой будет создана физическая и логическая структура хранилища данных:
  - • база данных - **farma.gdb**;
  - • логин - **sysdba**, пароль - **masterkey**;
  - • установить флажок **сохранять пароль**.
- 4. На следующем шаге выбирается версия для работы с ХД DeductorWarehouse
- 5. На пятом шаге при нажатии кнопки **Создать файл базы данных с необходимой структурой метаданных** по указанному ранее пути будет создан файл **farma.gdb** (появится сообщение об успешном создании). Это и есть пустое хранилище данных, готовое к работе.
- 6. На последних двух шагах осталось выбрать визуализатор для подключения (здесь это Сведения и Метаданные) и задать для нового хранилища имя **Farma**, метку **Фармация** и описание **«Хранилище данных с информацией о продажах в аптечной сети компании Фарма-Х»**.
- 6. После нажатия на кнопку **Готово** на дереве узлов подключений появится метка хранилища.

а

7. Для проверки доступа к новому ХД воспользуйтесь кнопкой — Если спустя некоторое время появится сообщение «Тестирование соединения прошло успешно», то хранилище готово к работе. Иначе нужно внести изменения в параметры подключения ХД, используя кнопку>—.

- 8. Сохраните настройки подключений, нажав на кнопку —.
- 9. Для проектирования структуры ХД вызвать **Редактор метаданных** кнопкой М на вкладке **Подключения**.

- 10. В открывшемся окне редактора, встав на узле **Измерения**, при помощи кнопки **Добавить** добавьте в метаданные первое измерение *Код группы* со следующими параметрами:
  - Имя - GR\_ID;
  - Метка - *Группа.Код*,
  - Тип данных - целый.
- 11. Прodelать аналогичные действия для создания всех остальных измерений, взяв параметры из таблицы:

| Измерение   | Имя     | Метка      | Тип данных |
|-------------|---------|------------|------------|
| Код группы  | GR.ID   | Группа.Код | целый      |
| Код товара  | TVJD    | Товар.Код  | целый      |
| Код отдела  | PART_ID | Отдел.Код  | целый      |
| Дата        | S_DATE  | Дата       | дата/время |
| Час покупки | S_HOUR  | Час        | целый      |

12. К каждому измерению, кроме *ДАТА* и *ЧАС*, нужно добавить по текстовому атрибуту. Для измерения *Группа.Код* это будет *Группа.Наименование* GR\_NAME, для измерения *Товар.Код* это будет *Товар.Наименование* TV\_NAME, для измерения *Отдел.Код* это будет *Отдел.Наименование* PART.NAME.

- 13. Путем простого добавления (ссылка на измерение отображается иконкой ®) установить ссылку измерения *Товар.Код* на измерение *Группа.Код*.
- 14. В окне редактора, встав на узле **Процессы**, при помощи кнопки **Добавить** введите в метаданные следующие параметры: Имя - SALES; Метка - *Продажи*.
- 15. Добавьте в процесс ссылки на 4 существующих измерения: *Дата*, *От-дел.Код*, *Товар.Код*, *Час* и добавьте ссылки на два факта: целочисленный - *Количество* F\_COUNT и вещественный - *Сумма* F\_SUM.
- 16. Проектирование структуры и метаданных ХД закончено, можно закрыть окно редактора.

## Задание 2. Загрузка информации в ХД

I. С помощью **Мастер импорта** последовательно импортируйте 4 текстовых файла в Deductor в следующей последовательности:

- groups.txt - товарные группы;
- produces.txt- товары;
- stores.txt - отделы;
- sales.txt - продажи товаров по дням.

*Замечание. В сценарии загрузки используйте относительные пути к текстовым файлам. При выборе файла для импорта используйте относительный путь, это означает, что файл должен находиться в той же папке, что и файл со сценарием.*

II. При загрузке данных в хранилище сначала загружаются таблицы измерений со своими атрибутами, и только после этого загружается таблица процесса sales.txt.

1. Для загрузки первого измерения *Группа.Код*, встав на первом узле сценария, вызовите **Мастер экспорта**, из списка приемников нужно выбрать DeductorWarehouse, далее из списка доступных хранилищ укажите ХД Фармация. На следующей вкладке требуется указать, в какое именно измерение будет загружаться

информация *Группа.Код*. Последнее, что осталось, это установить соответствие элементов объекта в ХД с полями вход-52

ного источника данных (т.е. таблицы groups.txt). Нажатие кнопки **Пуск** загружает в измерение данные. При этом старые данные, если они были, будут обновлены.

- 2. Прodelайте аналогичные действия еще для двух измерений - *Отдел.Код*, *Товар.Код*.
- 3. Загрузите данные в процесс *Продажи*, повторив сначала действия, аналогичные перечисленным выше. В отличие от загрузки измерений здесь появляются два специфических шага. На одном нужно задать параметры для контроля непротиворечивости информации в ХД - указать измерения, по которым следует удалять данные из хранилища (в нашем примере нужно поставить флажок по полю *Дата*). На последней странице настроек оставьте параметры по умолчанию.
- 4. Сохраните файл сценария под именем load.ded в той же папке, где находятся текстовые файлы таблиц.

### Задание 3. Извлечение информации из ХД

- 1. Импортируйте данные из процесса *Продажи* за последние три месяца.
- 1) С помощью **Мастера импорта** выберите тип источника данных DeductorWarehouse, на следующем шаге - ХД *Фармация*, а затем процесс - *Продажи*.
- 2) Определите, какие измерения и атрибуты из выбранного на предыдущем шаге процесса должны быть импортированы (поставьте флажки у следующих полей: *Дата*, *Отдел.Наименование*, *Товар.Наименование*, *Группа.Наименование*, *Час*).
- 3) Определите импортируемые факты (*Количество* и *Сумма*) и виды их агрегаций (в большинстве случаев требуется агрегация в виде суммы).
- 4) Для измерения *Дата* определите срез «Все продажи за последние три месяца от текущей даты» (нужно выделить это измерение, выбрать условие «последний» и значение «3 месяца от имеющихся данных»).
- 5) На следующем шаге настройте динамический фильтр по полю *Дата* (установите флажок).
- 6) Для результирующего набора данных определите способ его отображения в виде таблицы.
- 2. Импортируйте информацию о продажах из ХД, включая атрибуты товара. Установите следующие срезы:
  - • «кроме последнего периода, 1 месяц от имеющихся данных»;
  - • срез только по одной товарной группе.
- 3. Создайте новый сценарий farma.ded.
- 1) Создайте узлы импорта из ХД *Фармация*:

а) Список товарных групп.

б) Продажи по дням в Аптеке 1 по всем товарным группам, за исключением иммуномодуляторов за последние 10 недель от имеющихся данных.

2) Скопируйте последний узел и настройте в нем динамический фильтр на наименование отдела.

**Задание 4.** Требуется разработать систему аналитической отчетности в Deductor. Система отчетности строится на основе хранилища данных *Фармация*. Все требуемые отчеты должны быть вынесены на **Панель отчетов**.

- 1. **Диаграммы и гистограммы**
- 1) Постройте отчет - гистограмму распределения средних цен всех товаров, которые продавались за последние 5 месяцев от имеющихся данных. Назовите отчет «Гистограмма средних цен».
- 2) Постройте временной ряд продаж по месяцам, используя все имеющиеся данные (по оси ОУ откладывается сумма продаж).

- 2. OLAP-кубы
  - 1. Построить куб по трем измерениям (отдел, месяц года, товарная группа), в ячейках которого отображается сумма и объем (количество проданных единиц продукции) продаж за все периоды, имеющиеся в ХД.
  - 2. То же, что в п.1, но за последние три месяца от имеющихся данных.
  - 3. Сформировать многомерный отчет и график загруженности торговых точек по дням месяца.
  - 4. Сформировать многомерный отчет и график загруженности торговых точек по дням недели.
  - 5. 10 самых продаваемых товаров.
  - 6. 5 самых популярных товаров в каждой товарной группе.
  - 7. 10 самых продаваемых товаров с 18 до 21 часа.
  - 8. 10 самых продаваемых товаров по пятницам.
  - 9. 5 самых популярных товаров в товарной группе «Местные анестетики»
- Ю.Товары, дающие 80 % объема продаж в летние месяцы.
11. Пять товаров, пользующихся наибольшим спросом по понедельникам до 12 часов дня.

### **Вопросы к экзамену**

|                                                                                                                                                                                                                                |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Цель, задачи и функции бизнес-аналитики                                                                                                                                                                                        |
| Внедрение систем бизнес-аналитики на современных предприятиях: проблемы и перспективы                                                                                                                                          |
| Эволюция информационно-аналитических систем                                                                                                                                                                                    |
| Системы поддержки управленческих решений (DSS/BI)                                                                                                                                                                              |
| Системы бизнес-аналитики (БА)                                                                                                                                                                                                  |
| Профессиональные инструменты отчетности                                                                                                                                                                                        |
| Инструменты создания запросов и отчетов для конечных пользователей                                                                                                                                                             |
| Инструменты OLAP                                                                                                                                                                                                               |
| Инструменты оценочных и инструментальных панелей                                                                                                                                                                               |
| Инструменты «добычи данных» (Data Mining)                                                                                                                                                                                      |
| Инструменты планирования и моделирования                                                                                                                                                                                       |
| Классификация продуктов Businessintelligence                                                                                                                                                                                   |
| Аналитические приложения в корпоративных информационных системах                                                                                                                                                               |
| Платформы бизнес интеллекта                                                                                                                                                                                                    |
| Корпоративные BI-наборы (enterprise BI suites, EBIS)                                                                                                                                                                           |
| Системы визуализации данных и решений                                                                                                                                                                                          |
| Место и роль интеллектуального анализа данных (Data Mining, DM) в процессе принятия решений                                                                                                                                    |
| Анализ структурированной информации с помощью статистических и математических методов: отбор признаков, стратификация, кластеризация, ассоциации, визуализация, регрессия, прогнозирование временных рядов, последовательности |
| Анализ неструктурированной или слабоструктурированной информации: категоризация, разведка и семантическая обработка текстов, расширенный поиск информации и др.                                                                |
| Методики обнаружения нового знания в хранилищах данных (KDD)                                                                                                                                                                   |
| Процесс проектирования архитектуры BI-технологии                                                                                                                                                                               |
| Два слоя архитектуры BI-технологии: инфраструктура и прикладные сервисы                                                                                                                                                        |
| Концепция управления эффективностью бизнеса (Business Performance Management, BPM).                                                                                                                                            |
| Управление эффективностью бизнеса при помощи системы Project Expert .                                                                                                                                                          |
| Оценка эффективности систем бизнес-аналитики                                                                                                                                                                                   |

|                                                                                                                                                                                                                                |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Информационное обеспечение стратегического менеджмента                                                                                                                                                                         |
| Применение систем бизнес-аналитики в маркетинге                                                                                                                                                                                |
| Финансовый учет и корпоративная отчетность. Системы консолидации финансовой отчетности.                                                                                                                                        |
| Концепция управление эффективностью бизнеса (Business Performance Management, BPM) и ее основные элементы                                                                                                                      |
| Системы управления моделями (MMS)                                                                                                                                                                                              |
| Системы управления знаниями (KMS)                                                                                                                                                                                              |
| Оценка эффективности систем бизнес-аналитики                                                                                                                                                                                   |
| Основные тенденции развития систем бизнес-аналитики                                                                                                                                                                            |
| Обзор рынка BI-технологий, основные игроки на поле BI                                                                                                                                                                          |
| Цель, задачи и функции бизнес-аналитики                                                                                                                                                                                        |
| Внедрение систем бизнес-аналитики на современных предприятиях: проблемы и перспективы                                                                                                                                          |
| Эволюция информационно-аналитических систем                                                                                                                                                                                    |
| Системы поддержки управленческих решений (DSS/BI)                                                                                                                                                                              |
| Системы бизнес-аналитики (БА)                                                                                                                                                                                                  |
| Профессиональные инструменты отчетности                                                                                                                                                                                        |
| Инструменты создания запросов и отчетов для конечных пользователей                                                                                                                                                             |
| Инструменты OLAP                                                                                                                                                                                                               |
| Инструменты оценочных и инструментальных панелей                                                                                                                                                                               |
| Инструменты «добычи данных» (Data Mining)                                                                                                                                                                                      |
| Инструменты планирования и моделирования                                                                                                                                                                                       |
| Классификация продуктов Businessintelligence                                                                                                                                                                                   |
| Аналитические приложения в корпоративных информационных системах                                                                                                                                                               |
| Платформы бизнес интеллекта                                                                                                                                                                                                    |
| Корпоративные BI-наборы (enterprise BI suites, EBIS)                                                                                                                                                                           |
| Системы визуализации данных и решений                                                                                                                                                                                          |
| Место и роль интеллектуального анализа данных (Data Mining, DM) в процессе принятия решений                                                                                                                                    |
| Анализ структурированной информации с помощью статистических и математических методов: отбор признаков, стратификация, кластеризация, ассоциации, визуализация, регрессия, прогнозирование временных рядов, последовательности |
| Анализ неструктурированной или слабоструктурированной информации: категоризация, разведка и семантическая обработка текстов, расширенный поиск информации и др.                                                                |
| Методики обнаружения нового знания в хранилищах данных (KDD)                                                                                                                                                                   |
| Процесс проектирования архитектуры BI-технологии                                                                                                                                                                               |
| Два слоя архитектуры BI-технологии: инфраструктура и прикладные сервисы                                                                                                                                                        |
| Концепция управления эффективностью бизнеса (Business Performance Management, BPM).                                                                                                                                            |
| Основные понятия информационно-аналитических и интеллектуальных систем                                                                                                                                                         |
| Оценка эффективности систем бизнес-аналитики                                                                                                                                                                                   |
| Информационное обеспечение стратегического менеджмента                                                                                                                                                                         |
| Применение систем бизнес-аналитики в маркетинге                                                                                                                                                                                |
| Финансовый учет и корпоративная отчетность. Системы консолидации финансовой отчетности.                                                                                                                                        |
| Концепция управление эффективностью бизнеса (Business Performance Management, BPM) и ее основные элементы                                                                                                                      |
| Системы управления моделями (MMS)                                                                                                                                                                                              |

|                                                       |
|-------------------------------------------------------|
| Системы управления знаниями (KMS)                     |
| Оценка эффективности систем бизнес-аналитики          |
| Основные тенденции развития систем бизнес-аналитики   |
| Обзор рынка BI-технологий, основные игроки на поле BI |

## Список рекомендуемой литературы

### Основная литература:

1. Воронова Л.И., Воронов В.И. BigData. Методы и средства анализа: Учебное пособие / МТУСИ. – М., 2016. – 35 с.
2. Суляев, И. И. Визуализация систем управления Электронный ресурс / Суляев И. И. : практикум. - Норильск : НГИИ, 2017. - 176 с. - ISBN 978-5-89009-686-9, экземпляров неограничено
3. Сузи,, Р. А. Язык программирования Python : учебное пособие / Р. А. Сузи. - Язык программирования Python, 2022-07-28. - Электрон.дан. (1 файл). - Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. - 350 с. - электронный. - Книга находится в премиум-версии ЭБС IPRBOOKS. - ISBN 978-5-4497-0705-5, экземпляров неограничено

### Дополнительная литература:

1. Тягунов, Г.В. Математические методы анализа: учебное пособие для СПО / Е.А.Трофимова, С.В.Плотников, Д.В.Гилёв; под. ред. Е.А. Трофимовой. —Эл. изд.— 2-е изд., стер.—Саратов : Профобразование; Екатеринбург : Изд-во Урал. ун-та, 2019. — 271с.
2. Суляев, И. И. Визуализация систем управления Электронный ресурс / Суляев И. И. : практикум. - Норильск : НГИИ, 2017. - 176 с. - ISBN 978-5-89009-686-9, экземпляров неограничено
3. Учебно-методическое пособие по дисциплине Интеллектуальные информационные системы и технологии Электронный ресурс / сост. Е. Н. Турута. - Учебно-методическое пособие по дисциплине Интеллектуальные информационные системы и технологии, 2022-04-04. - Москва : Московский технический университет связи и информатики, 2014. - 24 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 2227-8397, экземпляров неограничено

### Интернет-ресурсы:

1. Официальный сайт библиотеки ФГАОУ ВО СКФУ Режим доступа: <http://catalog.ncstu.ru/catalog> –.
2. Информационная справочная система ГАРАНТ.РУ // Режим доступа: <http://www.garant.ru/>
3. Информационная справочная система КонсультантПлюс. // Режим доступа: <http://www.consultant.ru>
4. Инфраструктура научно-исследовательских данных — платформа доступа к данным для научных исследований // Режим доступа: <https://www.data-in.ru>
5. Профессиональное сообщество специалистов по обработке данных и машинному обучению// Режим доступа: <https://www.kaggle.com/>
6. Профессиональная база данных Росстата // Режим доступа: Росстат — Базы данных ([rosstat.gov.ru](http://rosstat.gov.ru))

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

# **МЕТОДИЧЕСКИЕ УКАЗАНИЯ**

по выполнению лабораторных работ

по дисциплине

«Информационно-аналитические системы и технологии»

направление 38.03.05 «Бизнес-информатика»

направленность (профиль) «Информационная бизнес-аналитика и цифровые  
технологии»

## **Введение**

Целью данной дисциплины является формирование у студента личностных и профессиональных качеств, позволяющих осуществлять профессиональную деятельность, связанную с анализом, внедрением и применением информационно-аналитических систем и технологий.

Задачами дисциплины «Информационно-аналитические системы и технологии» являются:

- изучение студентами использования методов автоматизации анализа информационной подготовки принятия управленческих решений с употреблением современных инструментальных средств широкого применения и специализированных пакетов прикладных программ;
- освоение основ разработки и сопровождения систем загрузки данных, информационных хранилищ, технологий оперативного и интеллектуального анализа данных, отражающих деятельность в различных предметных областях;
- познание основ проблематики и областей использования искусственного интеллекта, экспертных и основанных на знаниях систем.

## **Лабораторная работа № 1. Технологии анализа данных.**

**Цель работы.** Формирование знаний и навыков работы в среде интеллектуального анализа данных.

**Формируемые компетенции или их части:** ПК-4

### **Теоретическая часть**

Анализ данных — широкое понятие. Сегодня существуют десятки его определений. В самом общем смысле анализ данных — это исследования, связанные с обработкой многомерной системы данных, имеющей множество параметров. В процессе анализа данных исследователь производит совокупность действий с целью формирования определенных представлений о характере явления, описываемого этими данными. Как правило, для анализа данных используются различные математические методы.

Анализ данных нельзя рассматривать только как обработку информации после ее сбора. Анализ данных — это прежде всего средство проверки гипотез и решения задач исследователя.

Известное противоречие между ограниченными познавательными способностями человека и бесконечностью Вселенной заставляет нас использовать модели и моделирование, тем самым упрощая изучение интересующих объектов, явлений и систем.

Слово «модель» (лат. *modelium*) означает «мера», «способ», «сходство с какой-то вещью».

Построение моделей — универсальный способ изучения окружающего мира, позволяющий обнаруживать зависимости, прогнозировать, разбивать на группы и решать множество других задач. Основная цель моделирования в том, что модель должна достаточно хорошо отображать функционирование моделируемой системы.

АП «Deductor» применима для решения большого спектра задач, таких как создание аналитической отчетности, прогнозирование, поиск закономерностей и пр. Можно сказать, что данная система применима в задачах, где требуется консолидация и отображение данных различными способами, построение моделей и последующее применение полученных моделей к новым данным.

Рассмотрим некоторые задачи, решаемые АП:

- Системы корпоративной отчетности. Готовое хранилище данных и гибкие механизмы предобработки, очистки, загрузки, визуализации позволяют быстро создавать законченные системы отчетности в сжатые сроки.
- Обработка нерегламентированных запросов. Конечный пользователь может с легкостью получить ответ на вопросы типа "Сколько было продаж товара по группам в субъект за прошлый год с разбивкой по месяцам?" и просмотреть результаты наиболее удобным для него способом.
- Анализ тенденций и закономерностей, планирование, ранжирование. Простота использования и интуитивно понятная модель данных позволяет вам проводить анализ по принципу «Что, если...?», соотносить ваши гипотезы со сведениями, хранящимися в базе данных, находить аномальные значения, оценивать последствия принятия бизнес-решений.
- Прогнозирование. Построив модель на исторических примерах, вы можете использовать ее для прогнозирования ситуации в будущем. По мере изменения ситуации нет необходимости перестраивать все, необходимо всего лишь дообучить модель.
- Управление рисками. Реализованные в системе алгоритмы дают возможность достаточно точно определиться с тем, какие характеристики объектов и как влияют на риски, благодаря чему можно прогнозировать наступление рискованного события и заблаговременно принимать необходимые меры к снижению размера возможных неблагоприятных последствий.
- Анализ данных маркетинговых и социологических исследований. Анализируя сведения о потребителях, можно определить, кто является вашим клиентом и почему. Как

изменяются их пристрастия в зависимости от возраста, образования, социального положения, материального состояния и множества других показателей. Понимание этого будет способствовать правильному позиционированию ваших продуктов и стимулированию продаж.

- **Диагностика.** Механизмы анализа, имеющиеся в системе Deductor, с успехом применяются в медицинской диагностике и диагностике сложного оборудования. Например, можно построить модель на основе сведений об отказах. При ее помощи быстро локализовать проблемы и находить причины сбоев.

- **Обнаружение объектов на основе нечетких критериев.** Часто встречается ситуация, когда необходимо обнаружить объект, основываясь не на таких четких критериях, как стоимость, технические характеристики продукта, а на размытых формулировках, например, найти продукты, похожие на ваши с точки зрения потребителя.

#### **Задание.**

Изучить системы интеллектуального анализа данных.

Изучить интерфейс программы Deductor.

**Содержание отчета:** Отчёт по выполненной работе.

**Контрольные вопросы (перечень вопросов по теме, на которые студент обязан дать ответы) и /или тестовые задания**

1. Что такое методы ИАД и каково их назначение?
2. Охарактеризуйте области применения методов ИАД.
3. Каковы этапы исследований методами ИАД?

### **Лабораторная работа № 2. Принципы анализа данных**

**Цель работы.** Формирование знаний и навыков работы в среде интеллектуального анализа данных.

**Формируемые компетенции или их части:** ПК-4

#### **Теоретическая часть**

Процесс анализа

При информационном подходе к анализу данных, помимо модели, присутствуют еще три важные составляющие: эксперт, гипотеза и аналитик.

Эксперт — специалист в предметной области, профессионал, который за годы обучения и практической деятельности научился эффективно решать задачи, относящиеся к конкретной предметной области.

Эксперт — ключевая фигура в процессе анализа. По-настоящему эффективные аналитические решения можно получить не на основе одних лишь компьютерных программ, а в результате сочетания лучшего из того, что могут человек и компьютер. Эксперт выдвигает гипотезы (предположения) и для проверки их достоверности либо просматривает некие выборки различными способами, либо строит те или иные модели.

#### **ПРИМЕР-**

Гипотезой в анализе данных часто выступает предположение о влиянии какого-либо фактора или группы факторов на результат. К примеру, при построении прогноза продаж допускается предположение, что на величину будущих продаж существенно влияют продажи за предыдущие периоды и остатки на складе. При моделировании кредитоспособности потенциального заемщика выдвигается гипотеза, что на кредитоспособность влияют социально-экономические характеристики клиента: возраст, образование, семейное положение и т. п.

В крупных проектах по созданию прикладных аналитических решений участвуют, как правило, несколько экспертов, а также аналитик.

### Извлечение и визуализация данных

Чтобы получить новые знания об исследуемом объекте или явлении, не обязательно строить сложные модели. Часто достаточно «посмотреть» на данные в нужном виде, чтобы сделать определенные выводы или выдвинуть предположение о характере зависимостей в системе, получить ответ на интересующий вопрос. Это помогает сделать визуализация.

В случае визуализации аналитик некоторым образом формулирует запрос к системе, извлекает нужную информацию из различных источников и просматривает полученные результаты. На их основе он делает выводы, которые и являются результатом анализа. Существует множество способов визуализации данных: J O LAP (кросс-таблицы и кросс-диаграммы);

- таблицы;
- диаграммы, гистограммы;
- карты, проекции, срезы и т. п.

### Задание.

Провести сравнительный анализ систем визуализации данных.

**Содержание отчета:** Отчёт по выполненной работе.

**Контрольные вопросы (перечень вопросов по теме, на которые студент обязан знать ответы) и /или тестовые задания**

1. Что такое методы ИАД и каково их назначение?
2. Охарактеризуйте области применения методов ИАД.
3. Каковы этапы исследований методами ИАД?

### Лабораторная работа № 3. Извлечение из БД. Предварительная обработка

**Цель работы.** Формирование знаний и навыков работы в среде интеллектуального анализа данных.

**Формируемые компетенции или их части:** ПК-4

#### Теоретическая часть

Методы **Data Mining** входят в платформе Deductor в состав методов извлечения знаний из баз данных (**Knowledge Discovery in Databases**).

Процесс извлечения знаний выглядит следующим образом (рис.1).

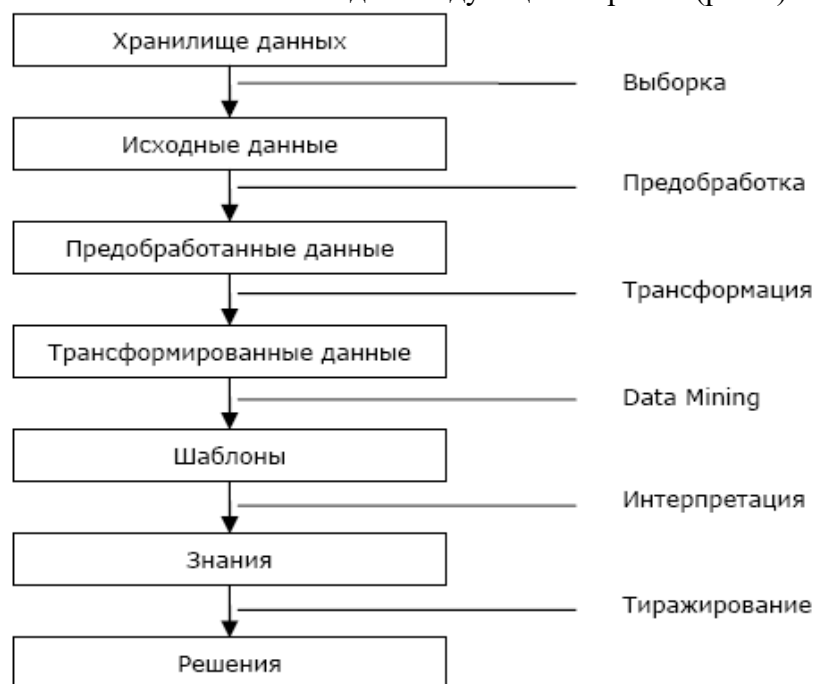


Рис.1

**DM – DataMining**– «разведка» данных. Это метод обнаружения в «сырых» данных ранее неизвестных, нетривиальных, практически полезных и доступных для интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. DM обеспечивает решение пяти задач — классификация, кластеризация, регрессия, ассоциация, последовательность.

1. *Классификация* — это отнесение объектов (наблюдений, событий) к одному из заранее известных классов. Классификация используется в случае, если заранее известны классы отнесения объектов. Например, отнесение нового товара к той или иной товарной группе, отнесение клиента к какой-либо категории. При кредитовании это может быть, например, отнесение клиента по каким-то признакам к одной из групп риска.

2. *Кластеризация* — это группировка объектов (наблюдений, событий) на основе данных (свойств), описывающих сущность объектов. Объекты внутри кластера должны быть «похожими» друг на друга и отличаться от объектов, вошедших в другие кластеры. Чем больше похожи объекты внутри кластера и чем больше отличий между кластерами, тем точнее кластеризация. Кластеризация может использоваться для сегментирования и построения профилей клиентов (покупателей). При достаточно большом количестве клиентов становится трудно подходить к каждому индивидуально. Поэтому клиентов удобно объединить в группы – сегменты со сходными признаками. Выделять сегменты клиентов можно по нескольким группам признаков. Это могут быть сегменты по сфере деятельности, по географическому расположению. После сегментации можно узнать, какие именно сегменты являются наиболее активными, какие приносят наибольшую прибыль, выделить характерные для них признаки. Эффективность работы с клиентами повышается за счет учета их персональных предпочтений.

3. *Регрессия* — установление функциональной зависимости между входными и непрерывными выходными переменными, в том числе для задачи прогнозирования. Регрессия чаще всего используется при прогнозировании объемов продаж, в этом случае зависимой величиной являются объемы продаж, а факторами, влияющими на эту величину, могут быть предыдущие объемы продаж, изменение курса валют, активность конкурентов и т.д. Или, например, при кредитовании физических лиц вероятность возврата кредита зависит от личных характеристик человека, сферы его деятельности, наличия имущества.

4. *Ассоциация* — выявление зависимостей между связанными событиями, указывающих, что из события X следует событие Y. Такие правила называются ассоциативными. Впервые эта задача была предложена для нахождения типичных шаблонов покупок, совершаемых в супермаркетах, поэтому иногда ее еще называют анализом потребительской корзины (marketbasketanalysis). Ассоциации помогают выявлять совместно приобретаемые товары. Это может быть полезно для более удобного размещения товара на прилавках, стимулирования продаж. Тогда человек, купивший пачку спагетти, не забудет купить к ним бутылочку соуса.

5. *Последовательные шаблоны* — установление закономерностей между связанными во времени событиями. Например, после события X через определенное время произойдет событие Y. Последовательные шаблоны могут быть использованы, например, при планировании продаж или предоставлении услуг. Например, если человек приобрел фото пленку, то через неделю он отдаст ее на проявку и закажет печать фотографий.

Иногда специально выделяют задачу *анализа отклонений* — выявление наиболее нехарактерных шаблонов. Для анализа отклонений необходимо сначала построить шаблон типичного поведения изучаемого объекта. Например, поведение человека при использовании кредитных карт. Тогда будет известно, что клиент (покупатель) использует карту регулярно два раза в месяц и приобретает товар в пределах определенной суммы. Отклонением будет, например, незапланированное приобретение товара по данной карте

на большую сумму. Это может говорить о ее использовании другим лицом, то есть о факте мошенничества.

Методика анализа с использованием механизмов Data Mining базируется на различных алгоритмах извлечения закономерностей из исходных данных, результатом работы которых являются модели. Таких алгоритмов довольно много, но, несмотря на их обилие, использование машинного обучения и т.п., они не способны гарантировать качественное решение. Никакой самый изощренный метод сам по себе не даст хороший результат, т.к. критически важным становится вопрос качества исходных данных. Чаще всего именно качество данных является причиной неудачи.

Ниже описана методика, следуя которой, можно подготовить качественные данные в нужном объеме для анализа. В этой последовательности действий все достаточно просто и логично, но, несмотря на это, пользователи почти всегда допускают одни и те же тривиальные ошибки.

Общая схема использования методов Data Mining состоит из следующих шагов (рис.2).



Рис.2

Эта последовательность действий не зависит от предметной области, поэтому ее можно использовать для любой сферы деятельности.

Теперь определим состав необходимых для решения проблемы методов.

- *Корреляционный анализ.* Позволяет проанализировать влияние входных факторов (выделенных на стадии выдвижения гипотезы) на результат и исключить незначащие факторы из дальнейшего анализа. Не имеет смысла использовать при анализе факторы, которые не вносят существенный вклад в результат и даже могут несколько исказить построенную модель.

- *OLAP.* Средства OLAP (OnLineAnalyticalProcessing) позволяют представить пользователю результаты обработки в удобной для восприятия и анализа форме. Как правило, наиболее удобными для анализа являются многомерные данные, описывающие предметную область сразу с нескольких точек зрения.

- *Кластерный анализ.* Разбиение всего массива налогоплательщиков на группы по сходным значениям факторов (параметров). Такое разбиение необходимо для определения для каждой из групп способа проведения проверки.

- *Классификация.* Отнесение новых рассматриваемых налогоплательщиков к заранее выделенным группам в зависимости от значений параметров организации-налогоплательщика.

- *Прогнозирование.* Построение прогноза возможного уровня прибыли от сбора налогов или от взимания штрафов с нарушителей.

## Задание.

### Импорт данных

Импорт данных является отправной точкой анализа данных. Импорт в Deductor может осуществляться из популярных форматов хранения данных, таких как Excel, Access, MS SQL, Oracle, Текстовый файл и прочих. Кроме того, имеется универсальный доступ к любому источнику данных посредством ADO или ODBC.

В налоговой инспекции используются таблицы в формате Text (текстовый файл с разделителями), из которого и будет осуществляться импорт в Deductor (рис.3).

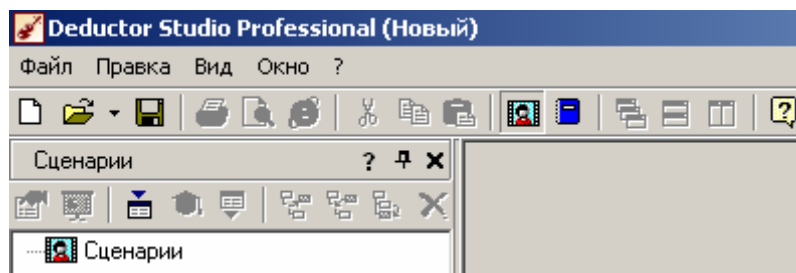


Рис.3

Импорт осуществляется путем вызова мастера импорта на панели «Сценарии» (рис.4).

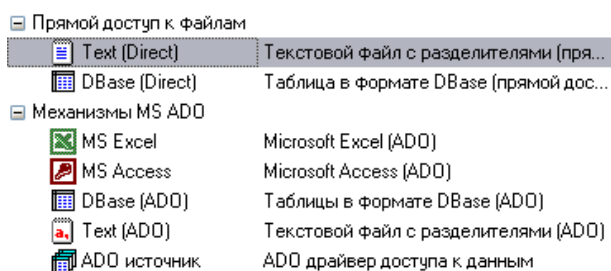


Рис.4

После запуска мастера импорта укажем тип импорта “Text (Direct)” и перейдем к настройке импорта. Укажем имя файла, из которого необходимо получить данные: "Data.txt" (рис.5).

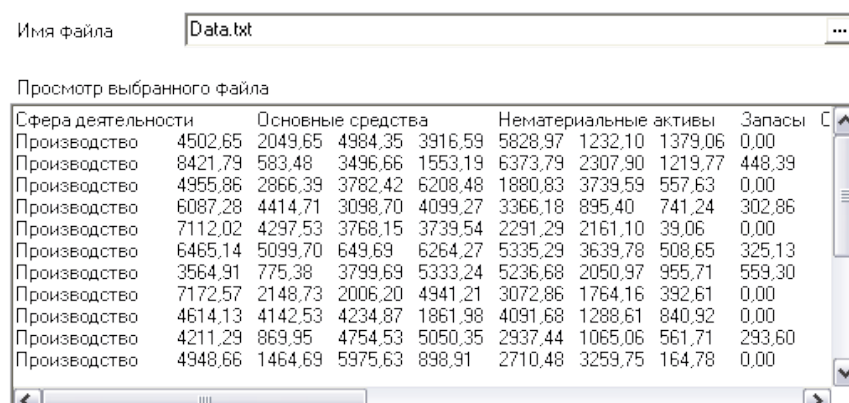


Рис.5

Далее предлагается настроить параметры импорта текстового файла, где указать символ-разделитель (символ табуляции) и обозначить то, что первая строка является заголовком (рис.6).

**Импорт текстового файла**  
Укажите параметры импорта текстового файла

Начать импорт со строки:  ☒ Первая строка является заголовком

Символом-разделителем является

☒ Символ табуляции ☐ Пробел ☐ Точка  
☐ Точка с запятой ☐ Запятая ☐ Другой

☐ Считать последовательные разделители одним

Ограничитель строк   
Разделитель целой и дробной частей числа   
Разделитель компонентов даты  Формат даты   
Разделитель компонентов времени  Формат времени

Рис.6

Далее перейдем к настройке свойств полей. На этом шаге мастера предоставляется возможность настроить имя, название (метку), размер, тип данных, вид данных и назначение. Некоторые свойства (например, тип данных) можно задавать для выделенного набора столбцов. Вид данных определяет – конечный это набор (дискретные) или бесконечный (непрерывные). Назначение столбцов определяет характер их использования в алгоритмах обработки (при импорте можно оставить значение по умолчанию) (рис.7).

|                                    |               |                                                 |
|------------------------------------|---------------|-------------------------------------------------|
| Сфера деятельности                 | Имя столбца   | <input type="text" value="COL1"/>               |
| 9.0 Основные средства              | Метка столбца | <input type="text" value="Сфера деятельности"/> |
| 9.0 Нематериальные активы          | Тип данных    | <input type="text" value="ab Строковый"/>       |
| 9.0 Запасы                         | Вид данных    | <input type="text" value="... Дискретный"/>     |
| 9.0 Отложенные налоговые активы    | Назначение    | <input type="text" value="i Информационное"/>   |
| 9.0 Доходы                         |               |                                                 |
| 9.0 Расходы                        |               |                                                 |
| 9.0 Уплаченные налог               |               |                                                 |
| 9.0 Максимальная дополнительная... |               |                                                 |

Рис.7

Далее проводится процесс загрузки данных.

После импорта данных на следующем шаге мастера необходимо выбрать способ отображения данных. В данном случае самым информативным является таблица, выберем ее (рис.8).

☒ Табличные данные

☒ Таблица ☐ Отображает данные в виде таблицы

☐ Статистика ☐ Отображает статистические данные выборки

☐ Диаграмма ☐ Отображает данные в виде диаграммы

☐ Гистограмма ☐ Отображает данные в виде гистограммы

☐ OLAP анализ

☐ Куб ☐ Многомерное отображение (кросс-таблица и кросс-диаграмма)

☐ Прочее

☐ Описание ☐ Сведения о параметрах

Рис.8

После всех действий сценарий будет иметь вид, представленный на рис.9.

Сценарии

Загрузка данных для построения модели

Рис.9

Для эффективного применения методов Data Mining следует обратить серьезное внимание на вопросы предобработки данных. Данные могут содержать пропуски, шумы, аномальные значения и т.д. Кроме того, данные могут быть избыточны, недостаточны и

т.д. В некоторых задачах требуется дополнить данные некоторой априорной информацией. Ошибочно предполагать, что если подать данные на вход системы в существующем виде, то на выходе будут получены полезные знания. Входные данные должны быть качественны и корректны.

В данном случае актуальной будет проверка на наличие пропусков. Часто бывает так, что в столбце некоторые данные отсутствуют в силу каких либо причин (данные не известны либо их забыли внести и т.п.). Обычно из-за этого пришлось бы убрать из обработки все строки, которые содержат пропущенные данные.

Для восстановления пропусков следует запустить мастер парциальной обработки (рис.10).

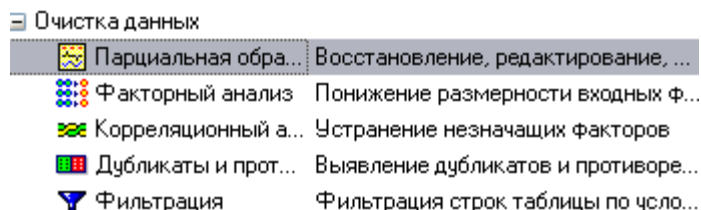


Рис.10

На следующем шаге необходимо выбрать тип обработки «Максимальное правдоподобие» для всех столбцов (рис.11).

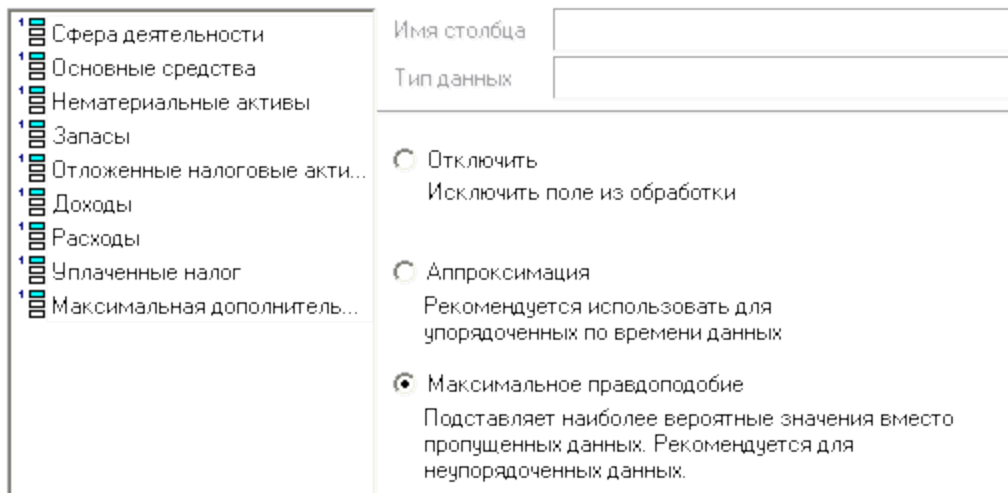


Рис.11

Остальные параметры не меняются, т.е. не будем проводить редактирование аномальных значений и спектральную обработку.

После выполнения процесса обработки пропуски в данных ликвидируются, что и было необходимо сделать.

Сценарий после добавления к загруженным данным обработчика "Парциальная обработка" будет выглядеть следующим образом (рис.12)

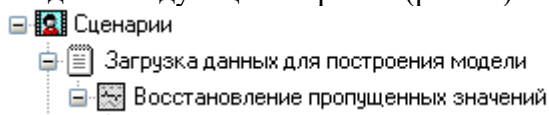


Рис.12

#### Задание.

Импортируйте текстовый файл с данными той же структуры "Task 1-2.txt" и выполните восстановление пропущенных значений. Укажите значения пропущенных ячеек в столбце «Основные средства» (записи номер 51 и 57), в столбце «Нематериальные активы» (запись номер 21), в столбце «Запасы» (записи номер 9 и 22). Сделайте выводы по работе алгоритма подстановки.

**Содержание отчета:** Отчёт по выполненной работе.

**Контрольные вопросы (перечень вопросов по теме, на которые студент обязан знать ответы) и /или тестовые задания**

1. Охарактеризуйте метод «Деревьев решений».
2. Охарактеризуйте особенности регрессионного анализа в методах ИАД.
3. Охарактеризуйте модели временных рядов с запаздываниями.
4. Охарактеризуйте метод «Ближайшего соседа».
5. Охарактеризуйте метод поиска правила.
6. Охарактеризуйте метод кластеризации.
7. Охарактеризуйте метод классификации.
8. Охарактеризуйте метод дискриминации.
9. Какие различия в целях и алгоритмах статистического и интеллектуального подходов.

#### **Лабораторная работа № 4. Корреляционный анализ.**

**Цель работы.** Формирование знаний и навыков работы в среде интеллектуального анализа данных.

**Формируемые компетенции или их части: ПК-4**

##### **Теоретическая часть**

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 незначущий.

Корреляция представляет собой меру зависимости переменных. Наиболее известна корреляция Пирсона. При вычислении корреляции Пирсона предполагается, что переменные измерены, как минимум, в интервальной шкале. Коэффициент корреляции Пирсона ( $r$ ) представляет собой меру линейной зависимости двух переменных. Коэффициенты корреляции изменяются в пределах от  $-1,00$  до  $+1,00$ . Значение  $-1,00$  означает, что переменные имеют строгую *отрицательную* корреляцию (две переменные могут быть связаны таким образом, что при возрастании значений одной из них значения другой убывают). Значение  $+1,00$  означает, что переменные имеют строгую *положительную* корреляцию (связь между двумя переменными может быть следующей – когда значения одной переменной возрастают, значения другой переменной также возрастают). Отметим, что значение  $0,00$  означает отсутствие корреляции.

Для проведения корреляционного анализа используем данные, подготовленные в лабораторной работе №1.

Добавляем к полученному сценарию новую ветку «Устранение незначущих факторов» после ветки «Восстановление пропущенных значений» (рис.13):

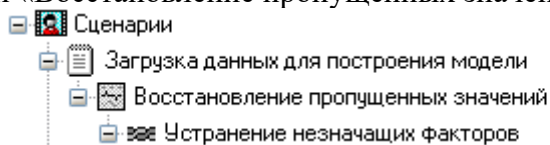


Рис.13

В мастере корреляционного анализа зададим все доходы, расходы и активы (основные средства, запасы и пр.) входными полями, а поле Возможный доход – выходным, поле Сфера деятельности будет непригодным, т.к. имеет строковый тип данных (рис.14).

Рис.14

Следующий шаг предлагает запустить процесс корреляционного анализа. После завершения процесса на следующем шаге 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). Включать в дальнейшую обработку будем те факторы, которые преодолели порог значимости в 0,5 (рис.15).

| Входные поля                                          | Корреляция с выходными полями     |        |
|-------------------------------------------------------|-----------------------------------|--------|
|                                                       | Максимальная дополнительная пр... |        |
| <input checked="" type="checkbox"/> Основные средс... |                                   | 0,507  |
| <input checked="" type="checkbox"/> Нематериальны...  |                                   | 0,508  |
| <input checked="" type="checkbox"/> Запасы            |                                   | 0,507  |
| <input type="checkbox"/> Отложенные нал...            |                                   | 0,288  |
| <input type="checkbox"/> Доходы                       |                                   | 0,214  |
| <input type="checkbox"/> Расходы                      |                                   | 0,289  |
| <input checked="" type="checkbox"/> Уплаченные налог  |                                   | -0,798 |

☐ Ручной выбор незначущих факторов  
☒ Автоматический выбор незначущих факторов в соответствии с порогом значимости

Порог значимости

Рис.15

Таким образом, корреляционный анализ позволил проанализировать влияние входных факторов на результат и исключить незначущие факторы из дальнейшего анализа.

### Задание.

Используя результаты обработки прошлого задания, проведите устранение незначущих факторов и укажите пороги значимости, а также включенные/исключенные из дальнейшей обработки факторы при пороге значимости 0,25.

**Содержание отчета:** Отчёт по выполненной работе.

**Контрольные вопросы (перечень вопросов по теме, на которые студент обязан знать ответы) и /или тестовые задания**

1. Охарактеризуйте метод «Деревьев решений».
2. Охарактеризуйте особенности регрессионного анализа в методах ИАД.
3. Охарактеризуйте модели временных рядов с запаздываниями.
4. Охарактеризуйте метод «Ближайшего соседа».
5. Охарактеризуйте метод поиска правила.

6. Охарактеризуйте метод кластеризации.
7. Охарактеризуйте метод классификации.
8. Охарактеризуйте метод дискриминации.
9. Какие различия в целях и алгоритмах статистического и интеллектуального подходов.

### **Лабораторная работа № 5. Кластерный анализ.**

**Цель работы.** Формирование знаний и навыков работы в среде интеллектуального анализа данных.

**Формируемые компетенции или их части: ПК-4**

#### **Теоретическая часть**

Самоорганизующаяся карта Кохонена является разновидностью нейронной сети. Она применяется, когда необходимо решить задачу кластеризации, т.е. распределить данные по нескольким кластерам. Алгоритм определяет расположение кластеров в многомерном пространстве факторов. Исходные данные будут относиться к какому-либо кластеру в зависимости от расстояния до него. Многомерное пространство трудно для представления в графическом виде. Механизм же построения карты Кохонена позволяет отобразить многомерное пространство в двумерном, которое более удобно и для визуализации и для интерпретации результатов аналитиком.

#### **Применение алгоритма**

Так как алгоритм SOM сочетает в себе два основных направления — векторное квантование и проецирование, то можно найти и основные применения этого алгоритма. Данную методику можно использовать для поиска и анализа закономерностей в исходных данных. При этом после того, как нейроны размещены на карте, полученная карта может быть отображена. Рассмотрим различные способы отображения полученной карты.

#### **Раскраска, порожденная отдельными компонентами**

При данном методе отрисовки полученную карту можно представить в виде слоеного пирога, каждый слой которого представляет собой раскраску, порожденную одной из компонент исходных данных. Полученный набор раскрасок может использоваться для анализа закономерностей, имеющихся между компонентами набора данных. После формирования карты мы получаем набор узлов, который можно отобразить в виде двумерной картинки. При этом каждому узлу карты можно поставить в соответствие участок на рисунке, четырех- или шестиугольный, координаты которого определяются координатами соответствующего узла в решетке. Теперь для визуализации осталось только определить цвет ячеек этой картинки. Для этого и используются значения компонент. Самый простой вариант — использование градаций серого. В этом случае ячейки, соответствующие узлам карты, в которые попали элементы с минимальными значениями компонента или не попало вообще ни одной записи, будут изображены черным цветом, а ячейки, в которые попали записи с максимальными значениями такого компонента, будут соответствовать ячейки белого цвета. В принципе, можно использовать любую градиентную палитру для раскраски.

Полученные раскраски в совокупности образуют атлас, отображающий расположение компонент, связи между ними, а также относительное расположение различных значений компонент.

#### **Отображение кластеров**

Кластером будет являться группа векторов, расстояние между которыми внутри этой группы меньше, чем расстояние до соседних групп. Структура кластеров при использовании алгоритма SOM может быть отображена путем визуализации расстояния между опорными векторами (весовыми коэффициентами нейронов). При использовании этого метода чаще всего используется унифицированная матрица расстояний (u-matrix),

вычисляется расстояние между вектором весов нейрона в сетке и его ближайшими соседями. Затем эти значения используются для определения цвета, которым этот узел будет отрисован. Обычно используют градации серого, причем чем больше расстояние, тем темнее отрисовывается узел. При таком использовании узлам с наибольшим расстоянием между ними и соседями соответствует черный цвет, а близлежащим узлам — белый.

### Реализация

Используются данные, подготовленные по результатам выполнения предыдущих лабораторных работ.

Добавляем к полученному сценарию новую ветку «Кластеризация», после ветки «Устранение незначущих факторов» (рис.16).

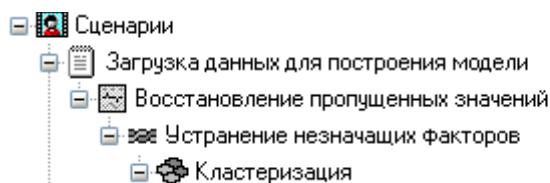


Рис.16

Запустим мастер обработки и выберем из списка метод обработки «Карта Кохонена». На втором шаге мастера настроим назначения столбцов. Укажем все столбцы входными (этот обработчик позволяет задать все поля входными и не указывать выходные), кроме поля «Максимальная дополнительная прибыль», которое зададим выходным. Т.е. предположим, что на основании значений поля «Максимальная дополнительная прибыль» организации могут быть отнесены к тому или иному классу.

На третьем шаге мастера необходимо настроить способ разделения исходного множества данных на тестовое и обучающее, а также количество примеров в том и другом множестве. Укажем, что данные обоих множеств берутся случайным образом, зададим размер тестового множества равным десяти примерам путем изменения значения столбца «Размер в строках» строки «Тестовое множество».

Следующий шаг предлагает настроить параметры карты (количество ячеек по X и по Y, их форму) и параметры обучения (способ начальной инициализации, тип функции соседства, необходимость перемешивания строк обучающего множества и количество эпох, через которые необходимо перемешивание). Значения по умолчанию вполне подходят (рис.17).

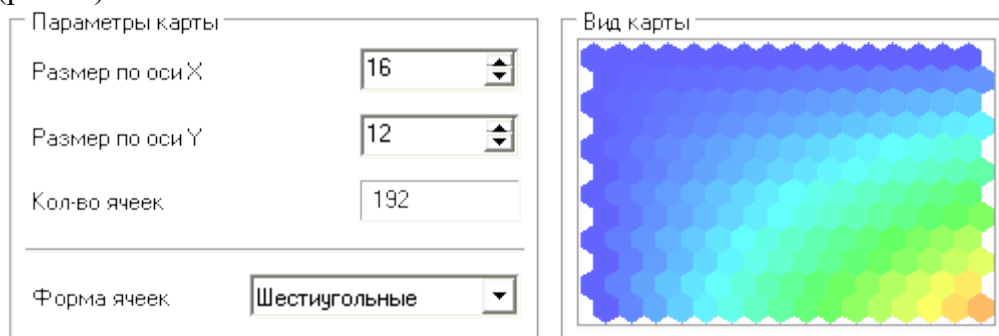


Рис.17

На пятом шаге мастера необходимо настроить параметры остановки обучения. Обучение будем останавливать при значении максимальной ошибки обучающего множества меньше 0,05 (рис.18).

Считать пример распознанным, если ошибка меньше

☐ По достижению эпохи

Обучающее множество

☐ Средняя ошибка меньше

☒ Максимальная ошибка меньше

☐ Распознано примеров (%)

Тестовое множество

☐ Средняя ошибка меньше

☐ Максимальная ошибка меньше

☐ Распознано примеров (%)

Рис.18

На шестом шаге настраиваются остальные параметры обучения – способ начальной инициализации, тип функции соседства и также параметры кластеризации – автоматическое определение числа кластеров с соответствующим уровнем значимости либо фиксированное количество кластеров. Предоставляется возможность настроить интервалы обучения. Каждый интервал задается количеством эпох, радиусом обучения и скоростью обучения. Укажем фиксированное количество кластеров, равное трем (рис.19).

Способ начальной инициализации карты

☒ Количество эпох, через которое необходимо перемешивать строки

Скорость обучения

В начале обучения

В конце обучения

Радиус обучения

В начале обучения

В конце обучения

Функция соседства

Кластеризация

☐ Автоматически определить количество кластеров

Уровень значимости, %  Фиксированное кол-во кластеров

Рис.19

На седьмом шаге предлагается запустить сам процесс обучения. Во время обучения можно посмотреть количество распознанных примеров и текущие значения ошибок.

Здесь необходимо нажать на кнопку пуск и дождаться завершения процесса обработки (рис.20).

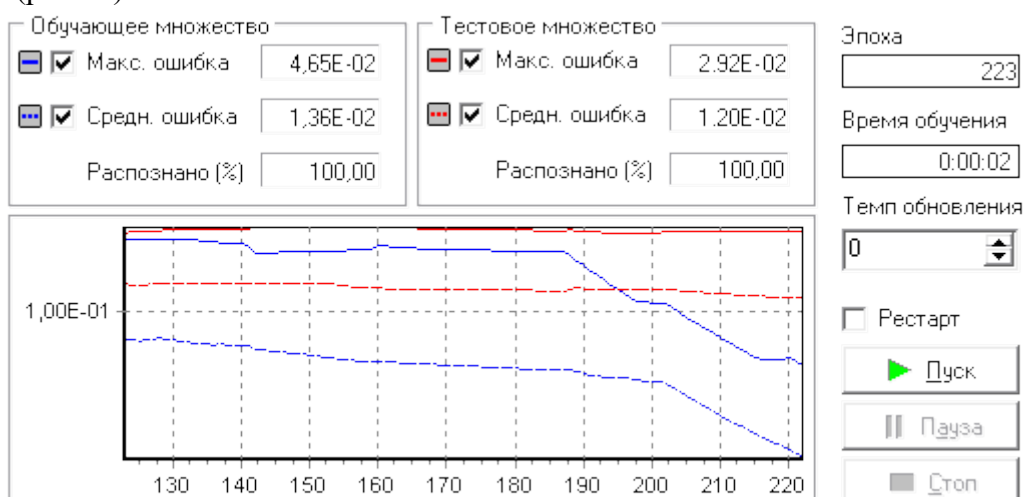


Рис.20

После этого необходимо в списке визуализаторов выбрать появившуюся теперь «Карту Кохонена» для просмотра результатов кластеризации, а также визуализатор «Что-если» для прогнозирования класса организации (проводить проверку или нет) (рис.21).

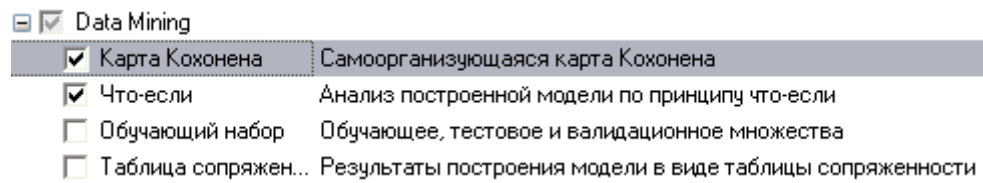


Рис.21

Далее в мастере настройки отображения карты Кохонена необходимо указать, чтобы отображались все поля, также следует поставить флажок «Границы кластеров» (рис.22).

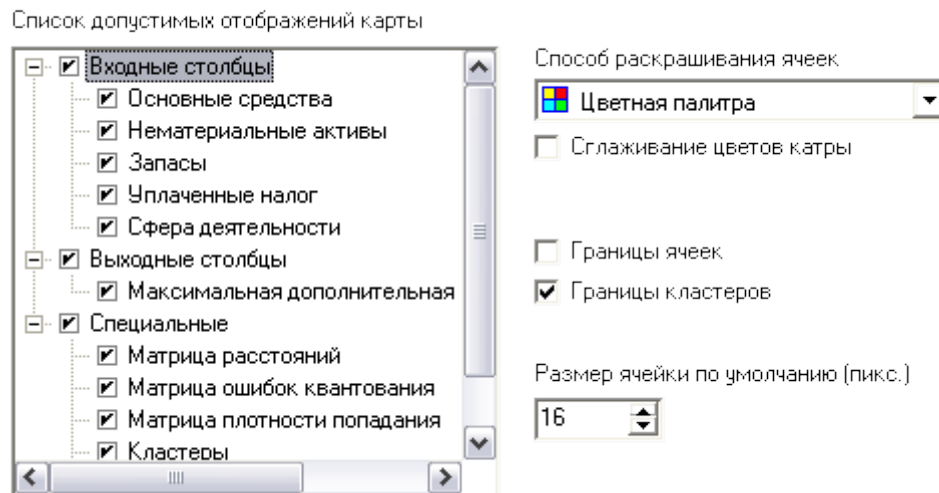


Рис.22

После этого можно увидеть полученные результаты обучения (рис.23).

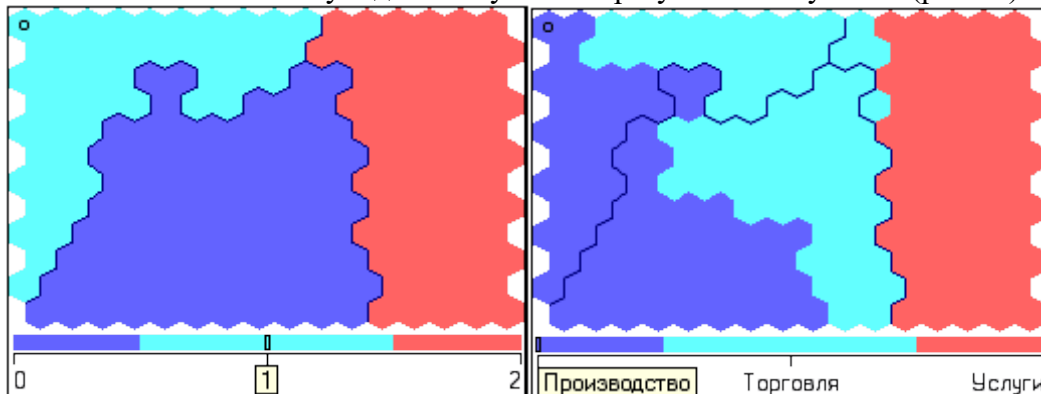


Рис.23

Как видно, все предприятия, работающие в сфере оказания услуг, относятся ко второму кластеру.

Кроме того, обработчик добавил новые поля, значения которых могут оказаться весьма полезными для эксперта, в дальнейшем использующего результаты кластеризации.

| Макс. дополнительная прибыль от проверки_OUT | Номер ячейки | Расстояние до центра ячейки | Номер кластера | Расстояние до центра кластера | Макс. дополнительная прибыль от проверки_ERR |
|----------------------------------------------|--------------|-----------------------------|----------------|-------------------------------|----------------------------------------------|
| 0                                            | 138          | 0,049075                    | 0              | 0,443099                      | 0                                            |
| 0                                            | 138          | 0,002234                    | 0              | 0,443099                      | 0                                            |

|        |     |          |   |          |   |
|--------|-----|----------|---|----------|---|
| 0      | 137 | 9,89E-07 | 0 | 0,363579 | 0 |
| 100,68 | 182 | 8,06E-06 | 0 | 0,352735 | 0 |
| 17,09  | 183 | 2,21E-06 | 0 | 0,505109 | 0 |
| 114,84 | 188 | 0,000155 | 2 | 0,538696 | 0 |
| 14,5   | 181 | 5,74E-06 | 0 | 0,461174 | 0 |
| 132,27 | 106 | 1,27E-05 | 0 | 0,383688 | 0 |

Полученные результаты кластеризации экспортируются во внешний текстовый файл. Далее эксперт проанализирует эти данные, добавит новый столбец «Проверка», где проставит вид проводимой проверки («Полная», «Не полная» или «Не проводить»).

#### **Задание.**

Классифицируйте организации со следующими параметрами.

| Сфера деятельности | Основные средства | Нематериальные активы | Запасы  | Уплаченный налог |
|--------------------|-------------------|-----------------------|---------|------------------|
| Производство       | 13281,17          | 4860,99               | 4171,59 | 2390,19          |
| Производство       | 8322,22           | 3347                  | 3026,8  | 1998,54          |
| Услуги             | 9274,75           | 921,98                | 2237,51 | 1492             |

**Содержание отчета:** Отчёт по выполненной работе.

**Контрольные вопросы (перечень вопросов по теме, на которые студент обязан знать ответы) и /или тестовые задания**

1. Охарактеризуйте метод «Деревьев решений».
2. Охарактеризуйте особенности регрессионного анализа в методах ИАД.
3. Охарактеризуйте модели временных рядов с запаздываниями.
4. Охарактеризуйте метод «Ближайшего соседа».
5. Охарактеризуйте метод поиска правила.
6. Охарактеризуйте метод кластеризации.
7. Охарактеризуйте метод классификации.
8. Охарактеризуйте метод дискриминации.
9. Какие различия в целях и алгоритмах статистического и интеллектуального подходов.

#### **Лабораторная работа № 6. Классификация с помощью деревьев решений**

**Цель работы.** Формирование знаний и навыков работы в среде интеллектуального анализа данных.

**Формируемые компетенции или их части: ПК-4**

#### **Теоретическая часть**

Деревья решений (decision trees) являются одним из наиболее популярных подходов к решению задач добычи данных. Они создают иерархическую структуру классифицирующих правил типа «ЕСЛИ...ТО...» (if-then), имеющую вид дерева. Чтобы принять решение, к какому классу следует отнести некоторый объект или ситуацию, требуется ответить на вопросы, стоящие в узлах этого дерева, начиная с его корня. Вопросы имеют вид «значение параметра А больше В?». Если ответ положительный,

осуществляется переход к правому узлу следующего уровня. Затем снова следует вопрос, связанный с соответствующим узлом и т. д. В каждом узле бинарных деревьев ветвление производится по двум направлениям (т. е. на вопрос, заданный в узле, имеется только два варианта ответов, например, Да или Нет). Однако в общем случае ответов, а следовательно, и ветвей, выходящих из узла, может быть больше.

Дерево решений состоит из узлов – где производится проверка условия, и листьев – конечных узлов дерева, указывающих на класс (узлов решения).

Также после построения присутствует информация о достоверности того или иного правила, его значимость. С помощью данного инструмента можно узнать ранг значимости каждого фактора (наиболее значимые факторы находятся на верхних уровнях дерева).

Полученную обучающую выборку (сформированную экспертом) загружают в систему (текстовый файл " training\_sample.txt").

Эксперт формирует таблицу из результатов кластеризации путём оставления столбцов Основные средства, Нематериальные активы, Запасы, Уплаченный налог и сфера деятельности, и добавления к ним столбца Проверка, в который он заносит свои оценки назначаемой проверки.

Добавляем к корню дерева сценария новую ветку «Загрузка экспертных данных» и после неё ветку «Построение решений» (рис.24):

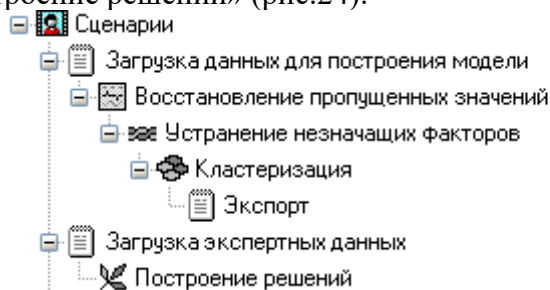


Рис.24

Выберем в качестве обработки дерево решений. В мастере построения дерева решения на втором шаге настроим «Название организации» как информационный, «Проверка» – как выходной, остальные поля – входные. Далее предлагается настроить способ разбиения исходного множества данных на обучающее и тестовое. Зададим способ разбиения по порядку, когда данные для тестового и обучающего множества берутся из исходного набора подряд, т.е. для обучающего множества будут использоваться проанализированные экспертом данные. На следующем шаге мастера предлагается настроить параметры процесса обучения, а именно минимальное количество примеров, при котором будет создан новый узел (пусть узел создается, если в него попали два и более примеров), а также предлагается возможность строить дерево с более достоверными правилами и отсекают узлы дерева. Включим данные опции (рис.25).

Параметры ранней остановки

Минимальное количество примеров в узле, при котором будет создан новый
2

☒ Строить дерево с более достоверными правилами в ущерб компактности дерева  
Очередной узел будет разбиваться на подузлы, если количество нераспознаваемых примеров в узле больше значения параметра "минимальное количество примеров в узле".

Параметры отсечения

☒ Отсекать узлы дерева  
Уровень доверия используемый при отсечении узлов дерева, %
20,0

Чем меньше уровень доверия, тем больше узлов будет отсечено и тем компактнее будет дерево.

Рис.25

На следующем шаге мастера запускается сам процесс построения дерева. Также можно увидеть информацию о количестве распознанных примеров (рис.26).

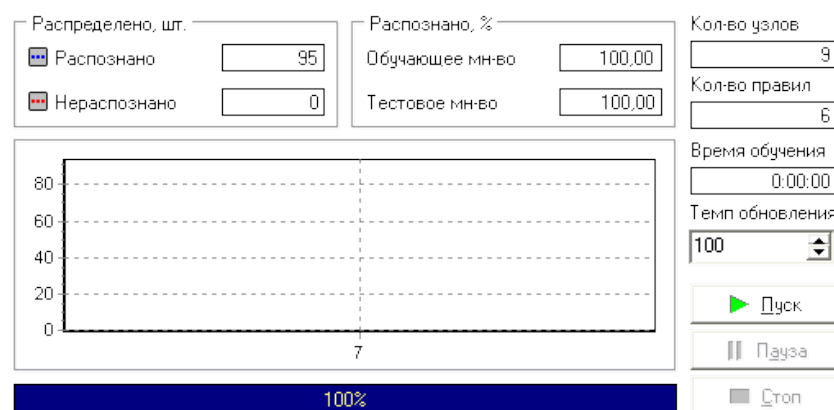


Рис.26

Получили следующее дерево решений и правила в соответствии с ним (рис.27).

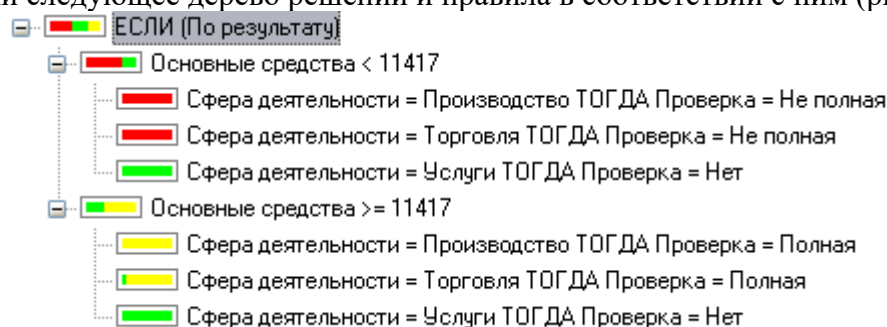


Рис.27

| N | Условие                                                        | Следствие (Проверка) | Поддержка |        | Достоверность |        |
|---|----------------------------------------------------------------|----------------------|-----------|--------|---------------|--------|
|   |                                                                |                      | %         | Кол-во | %             | Кол-во |
| 1 | Основные средства < 11417 И Сфера деятельности = Производство  | Не полная            | 24,21     | 23     | 100,00        | 23     |
| 2 | Основные средства < 11417 И Сфера деятельности = Торговля      | Не полная            | 20,00     | 19     | 100,00        | 19     |
| 3 | Основные средства < 11417 И Сфера деятельности = Услуги        | Нет                  | 16,84     | 16     | 100,00        | 16     |
| 4 | Основные средства >= 11417 И Сфера деятельности = Производство | Полная               | 12,63     | 12     | 100,00        | 12     |
| 5 | Основные средства >= 11417 И Сфера деятельности = Торговля     | Полная               | 12,63     | 12     | 100,00        | 12     |
| 6 | Основные средства >= 11417 И Сфера деятельности = Услуги       | Нет                  | 13,68     | 13     | 100,00        | 13     |

Рис.28

С помощью визуализатора "Что-если" можно протестировать модель на других данных, подставляя соответствующие значения (рис.29).

|                       |                  |
|-----------------------|------------------|
| Входные               |                  |
| 9.0 Основные средс... | 1547,84          |
| 9.0 Нематериальны...  | 35,65            |
| 9.0 Запасы            | 309,57           |
| 9.0 Уплаченные налог  | 1734,22          |
| ab Сфера деятельн...  | Производство     |
| Выходные              |                  |
| ab Проверка           | Не полная        |
| Расчетные             |                  |
| 12 Проверка Номер...  | 1                |
| 9.0 Проверка Подде... | 24,2105263157895 |
| 9.0 Проверка Досто... | 100              |

Рис.29

### Задание.

Определите вид проверки для следующих организаций.

| Название организации | Сфера деятельности | Основные средства | Нематериальные активы | Запасы  | Уплаченный налог |
|----------------------|--------------------|-------------------|-----------------------|---------|------------------|
| Организация 1        | Производство       | 12202             | 81,4                  | 387,6   | 540              |
| Организация 2        | Торговля           | 1287,59           | 113,72                | 257,52  | 679              |
| Организация 3        | Услуги             | 9914              | 6063                  | 4375,65 | 1533             |

**Содержание отчета:** Отчёт по выполненной работе.

**Контрольные вопросы (перечень вопросов по теме, на которые студент обязан знать ответы) и /или тестовые задания**

1. Охарактеризуйте генетические алгоритмы.
2. Охарактеризуйте нейросетевые методы анализа.
3. Охарактеризуйте методы для анализа нечетких множеств.
4. Перечислите основные направления эволюционного моделирования и приведите основные факторы, определяющие неизбежность эволюции.
5. В чем особенности эволюционного программирования? Приведите основные шаги обобщенного алгоритма эволюционного программирования.
6. Охарактеризуйте метод эволюционных стратегий. В чем его отличие от эволюционного программирования и от генетических алгоритмов?
7. Применение эволюционных вычислений в ИИС.
8. Какие алгоритмы называют генетическими? Сформулируйте основные особенности генетических алгоритмов.
9. Охарактеризуйте простой генетический алгоритм. Приведите пример.

### Лабораторная работа № 7. OLAP.

**Цель работы.** Формирование знаний и навыков работы в среде интеллектуального анализа данных.

**Формируемые компетенции или их части:** ПК-4

### Теоретическая часть

**OLAP** (On-LineAnalyticalProcessing) является ключевым компонентом организации хранилищ данных. Эта технология основана на построении и визуализации многомерных кубов данных с возможностью произвольного манипулирования информацией, содержащейся в кубе. Это позволяет представить информацию для анализа в любом разрезе.

Для решения задачи многомерного представления данных также весьма полезны сведения по объемам сбора того или иного вида налога за некоторый период, к примеру, за только что завершившийся отчетный период. Организации-налогоплательщики работают не в одной сфере деятельности, а в различных (производство, оказание услуг и т.д.). Первые два простейших вопроса, на которые нам сразу же хотелось бы иметь ответы, – это объемы сбора налогов (конкретных видов) по отдельным сферам деятельности за отчетный период.

Очевидно, что «ответ» на каждый из этих вопросов будет оформлен в виде двумерной таблицы. В первом случае строками и столбцами этой таблицы соответственно будут названия налога, период и суммы, а во втором – названия сферы деятельности и суммы.

Однако анализировать информацию в таком виде неудобно. Возникает потребность «соединить» данные нескольких таблиц (представим их в виде трехмерного куба). В итоге в таком отчете будет фигурировать три равноправных аналитических измерения (город, товар и месяц), и вместо двумерных (рис.30).

*Рис.30.*

Данные для обработки находятся в файле "OLAPData.txt". Импортируем данные и в визуализаторах выбираем "OLAP анализ"/"Куб".

При многомерном представлении данных вся информация представляется структурой типа «звезда», где в центре расположены таблицы фактов, а «лучами» являются измерения. Измерения могут быть как простыми списками, например, дата (Период), так и содержать дополнительные столбцы, называемые свойствами (рис.31).

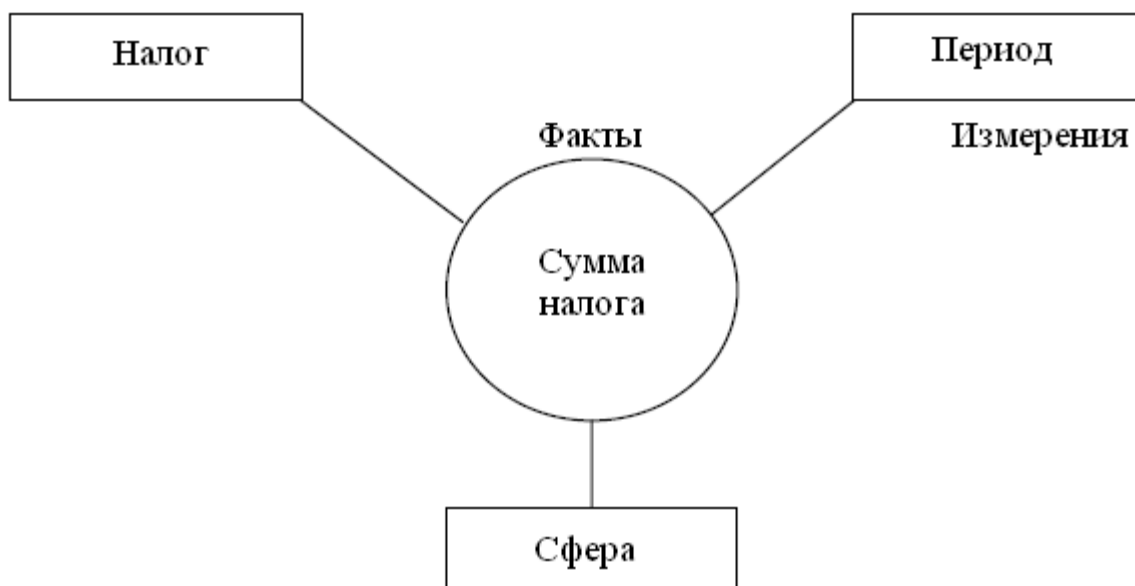


Рис.31

В соответствии с этим назначаем поля Период, Вид налога, Сфера деятельности измерениями, а поле Уплаченные налог – фактом.

|                    |             |             |
|--------------------|-------------|-------------|
| Сфера деятельности | Имя столбца | COL3        |
| Уплаченные налог   | Тип данных  | Дата/Время  |
| Период             | Назначение  | Измерение   |
| Вид налога         | Вид данных  | Непрерывный |

Статистическая информация недоступна.

Рис.32

Измерения могут быть размещены в строках и столбцах кросс-таблицы. В мастере настройки отображения изначально весь список выбранных измерений отображается в окне «Доступные измерения». Нажимая кнопки «>» и «>>» справа и снизу от этого окна, можно размещать выбранные измерения в строках и столбцах таблицы. Один из способов размещения приведен на рис.33.

|                     |    |                                                      |           |
|---------------------|----|------------------------------------------------------|-----------|
| Доступные измерения | >  | Измерения в столбцах                                 |           |
|                     | >> | Вид налога                                           |           |
|                     | <  |                                                      |           |
|                     | << |                                                      |           |
| Измерения в строках |    | Факт                                                 | Агрегация |
| Период              |    | <input checked="" type="checkbox"/> Уплаченные налог | Σ Сумма   |
| Сфера деятельности  |    |                                                      |           |

Рис.33

Здесь можно также выбрать, какие факты отображать в кросс-таблице на пересечении измерений и какую функцию применять при их агрегации (объединении). В данном примере факт «Уплаченный налог» будет суммироваться.

В результате получим кросс-таблицу, по которой удобно анализировать информацию, например, наибольшая сумма Налога 1 была собрана в 2002 г., причём наибольший вклад в эту сумму был с предприятий, работающих в сфере торговли (рис.34).

|        |                    | Вид налога |           |           |           |           |           |           |
|--------|--------------------|------------|-----------|-----------|-----------|-----------|-----------|-----------|
| Период | Сфера деятельности | Налог 1    | Налог 10  | Налог 2   | Налог 3   | Налог 4   | Налог 5   | Налог 6   |
| 2000   | Производство       | 4 249,42   | 2 283,25  | 1 805,79  | 5 610,64  | 8 191,18  | 7 995,88  | 9 332,83  |
|        | Торговля           | 4 182,52   | 3 362,73  | 4 074,40  | 8 356,60  | 4 786,79  | 4 534,61  | 3 716,98  |
|        | Услуги             | 3 274,49   | 4 086,27  | 773,92    | 382,72    | 2 555,93  | 3 066,23  | 294,83    |
|        | Итого              | 11 706,43  | 9 732,25  | 6 654,11  | 14 349,96 | 15 533,90 | 15 596,72 | 13 344,64 |
| 2001   | Производство       | 9 153,64   | 6 894,03  | 969,01    | 6 347,37  | 7 255,48  | 9 797,29  | 4 788,71  |
|        | Торговля           | 1 443,95   | 2 210,65  | 8 619,18  | 1 800,59  | 3 366,05  | 6 894,28  | 6 166,78  |
|        | Услуги             | 1 389,53   | 3 158,81  | 2 881,91  | 1 196,07  | 4 092,81  | 4 918,80  | 5 070,29  |
|        | Итого              | 11 987,12  | 12 263,49 | 12 470,10 | 9 344,03  | 14 714,34 | 21 610,37 | 16 025,78 |
| 2002   | Производство       | 6 734,99   | 6 218,00  | 5 806,59  | 2 650,41  | 4 397,63  | 5 328,84  | 6 394,49  |
|        | Торговля           | 10 514,22  | 7 070,83  | 8 609,39  | 8 259,76  | 5 282,40  | 5 004,88  | 6 134,25  |
|        | Услуги             | 5 180,60   | 6 596,02  | 7 074,60  | 7 155,02  | 3 408,67  | 2 692,02  | 7 777,97  |
|        | Итого              | 22 429,81  | 19 884,85 | 21 490,58 | 18 065,19 | 13 088,70 | 13 025,74 | 20 306,71 |
| 2003   | Производство       | 6 110,56   | 3 866,60  | 6 544,77  | 5 533,68  | 2 432,67  | 8 095,49  | 4 221,97  |
|        | Торговля           | 3 626,01   | 4 161,11  | 7 123,30  | 7 007,61  | 6 141,04  | 6 728,42  | 9 076,56  |
|        | Услуги             | 3 690,51   | 3 364,86  | 7 123,57  | 6 283,02  | 3 479,24  | 8 262,45  | 597,66    |
|        | Итого              | 13 427,08  | 11 392,57 | 20 791,64 | 18 824,31 | 12 052,95 | 23 086,36 | 13 896,19 |
| 2004   | Производство       | 7 595,69   | 4 208,91  | 5 358,76  | 7 852,60  | 4 629,25  | 3 050,46  | 7 689,98  |
|        | Торговля           | 5 469,68   | 6 070,39  | 6 354,89  | 4 375,95  | 1 947,28  | 8 166,71  | 4 277,78  |
|        | Услуги             | 3 898,08   | 8 330,41  | 1 585,02  | 9 982,39  | 3 583,78  | 5 215,32  | 4 414,35  |
|        | Итого              | 16 963,45  | 18 609,71 | 13 298,67 | 22 210,94 | 10 160,31 | 16 432,49 | 16 382,11 |
| Итого  |                    | 76 513,89  | 71 882,87 | 74 705,10 | 82 794,43 | 65 550,20 | 89 751,68 | 79 955,43 |

Рис.34

#### Задание.

Измените куб и интерпретируйте полученные результаты.

**Содержание отчета:** Отчёт по выполненной работе.

**Контрольные вопросы (перечень вопросов по теме, на которые студент обязан знать ответы) и /или тестовые задания**

1. Опишите операторы репродукции и кроссинговера в простом генетическом алгоритме. Приведите примеры.
2. Фундаментальная теорема генетического алгоритма.
3. Приведите пример применения фундаментальной теоремы генетического алгоритма.

#### Лабораторная работа № 8. Прогнозирование.

**Цель работы.** Формирование знаний и навыков работы в среде интеллектуального анализа данных.

**Формируемые компетенции или их части: ПК-4**

##### Теоретическая часть

Прогнозирование позволяет получать предсказание значений временного ряда на число отсчетов, соответствующее заданному горизонту прогнозирования. Алгоритм прогнозирования работает следующим образом. Пусть в результате преобразования методом скользящего окна была получена последовательность временных отсчетов:

$X(-n), \dots, X(-2), X(-1), X,$   
где  $X$  – текущее значение.

Прогноз на  $X(+1)$  строится на основании построенной модели. Чтобы построить прогноз для значения  $X(+2)$ , нужно сдвинуть всю последовательность на один отсчет влево, чтобы ранее сделанный прогноз  $X(+1)$  тоже вошел в число исходных значений.

Затем снова будет запущен алгоритм расчета прогнозируемого значения –  $X(+2)$  будет рассчитан с учетом  $X(+1)$  и так далее в соответствии с заданным горизонтом прогноза.

Для настройки алгоритма прогнозирования необходимо задать горизонт прогноза, а также поля таблицы, которые необходимо подавать на вход модели для построения прогноза (для вычисления выходного поля модели).

Проведем прогнозирование количества должников по месяцам с помощью нейросети.

Данные для прогнозирования находятся в файле "Prognoz.txt" и содержат статистику количества должников по месяцам за предыдущие годы. Импортируем их (рис.35).

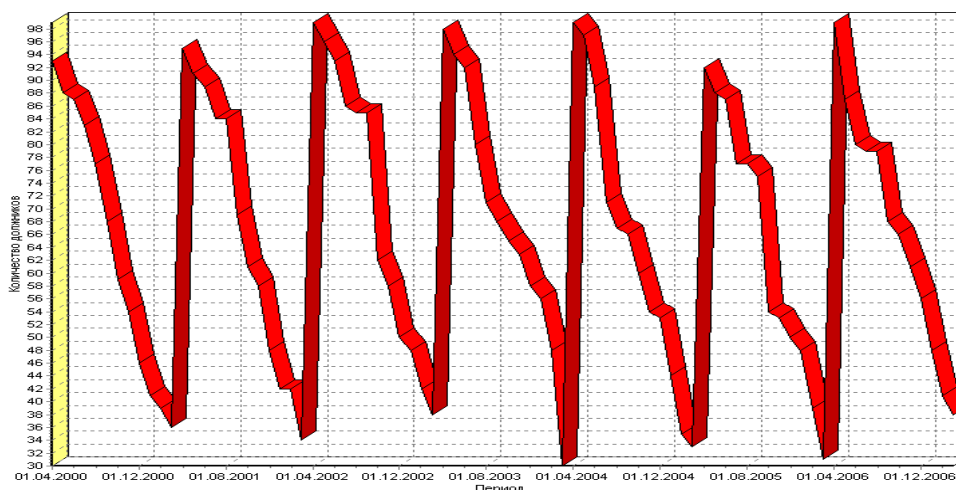


Рис.35

Для дальнейшего прогнозирования необходимо соответствующим образом подготовить данные. При прогнозировании временных рядов при помощи нейросети требуется подавать на вход анализатора значения нескольких, смежных, отсчетов из исходного набора данных. Такой метод отбора данных называется скользящим окном (окно – поскольку выделяется только некоторый непрерывный участок данных, скользящее – поскольку это окно «перемещается» по всему набору). При этом эффективность реализации заметно повышается, если не выбирать данные каждый раз из нескольких последовательных записей, а последовательно расположить данные, относящиеся к конкретной позиции окна, в одной записи.

Значения в одном из полей записи будут относиться к текущему отсчету, а в других – смещены от текущего отсчета «в будущее» или «в прошлое». Таким образом, преобразование скользящего окна имеет два параметра: «глубина погружения» – количество «прошлых» отсчетов, попадающих в окно, и «горизонт прогнозирования» – количество «будущих» отсчетов. Следует отметить, что для граничных (относительно начала и конца всей выборки) положений окна будут формироваться неполные записи, т.е. записи, содержащие пустые значения для отсутствующих прошлых или будущих отсчетов. Алгоритм преобразования позволяет исключить такие записи из выборки (тогда для нескольких граничных отсчетов записи формироваться не будут) либо включить их (тогда формируются записи для всех имеющихся отсчетов, но некоторые из них будут неполными). Отметим, что для правильного формирования скользящего окна данные должны быть соответствующим образом упорядочены.

В нашем случае имеются данные по количествам должников по месяцам за 7 лет, значит, выбираем поле "Количество должников" как используемое и выбираем глубину погружения 12, т.к. в году 12 месяцев (рис.36).

Имя столбца: COL2

Тип данных: Вещественный

Назначение: ☒ Используемое

Глубина погружения: 12

Горизонт прогнозирования: 0

☐ Оставлять неполные записи

Рис.36

Далее предстоит построить модель, для чего в качестве обработчика выбираем "Нейросеть". Поле "Период" делаем информационным, все поля, сформированные скользящим окном (Количество должников-1 – Количество должников-12), – входными, а поле "Количество должников" – выходным (рис.37).

Имя столбца: COL1

Тип данных: Дата/Время

Назначение: ☒ Информационное

Вид данных: Непрерывный

Статистическая информация недоступна.

Настройка нормализации...

Рис.37

Далее назначаем количество элементов в тестовом и обучающем множестве и выбираем структуру нейронной сети. Оставим параметры по умолчанию (рис.38).

Нейроны в слоях

входном: 12

скрытых слоев: 1

выходном: 1

| Слой | Нейроны |
|------|---------|
| 1    | 2       |

Активационная функция

Тип функции: Сигмоида

Крутизна: 1,000

Сигмоида

График функции Сигмоида: y-ось от 0,0 до 1,0, x-ось от -10 до 10.

Рис.38

Далее можно выбрать алгоритм и параметры обучения, оставляем и их по умолчанию (рис.39).

| Алгоритм                                                                                                                                                                                                                              | Параметры                                                                                                                                                            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input type="radio"/> Back - Propagation<br>Обучение в режиме "онлайн". Коррекция весов производится после предъявления каждого примера обучающего множества.                                                                         | Шаг спуска <input type="text" value="0,5"/><br>В случае изменения знака градиентной составляющей ошибки для данного веса задает величину следующей коррекции веса.   |
| <input checked="" type="radio"/> Resilient Propagation (RPROP)<br>Обучение в режиме "оффлайн". Коррекция весов производится после предъявления всех примеров обучающего множества. Учитывается только знак градиента по каждому весу. | Шаг подъема <input type="text" value="1,2"/><br>В случае сохранения знака градиентной составляющей ошибки для данного веса задает величину следующей коррекции веса. |

Рис.39

На следующем шаге нужно настроить параметры остановки обучения, где выберем остановку обучения, если максимальная ошибка обучающего множества меньше 0,05, т.е. при данной ошибке качество обучения будет достаточно приемлемым (рис.40).

|                                                                |                                    |
|----------------------------------------------------------------|------------------------------------|
| Считать пример распознанным, если ошибка меньше                | <input type="text" value="0,05"/>  |
| <input type="checkbox"/> По достижению эпохи                   | <input type="text" value="10000"/> |
| Обучающее множество                                            |                                    |
| <input type="checkbox"/> Средняя ошибка меньше                 | <input type="text"/>               |
| <input checked="" type="checkbox"/> Максимальная ошибка меньше | <input type="text" value="0,05"/>  |
| <input type="checkbox"/> Распознано примеров (%)               | <input type="text" value="0"/>     |
| Тестовое множество                                             |                                    |
| <input type="checkbox"/> Средняя ошибка меньше                 | <input type="text"/>               |
| <input type="checkbox"/> Максимальная ошибка меньше            | <input type="text"/>               |
| <input type="checkbox"/> Распознано примеров (%)               | <input type="text" value="0"/>     |

Рис.40

Запускаем обучение нейронной сети (рис.41).

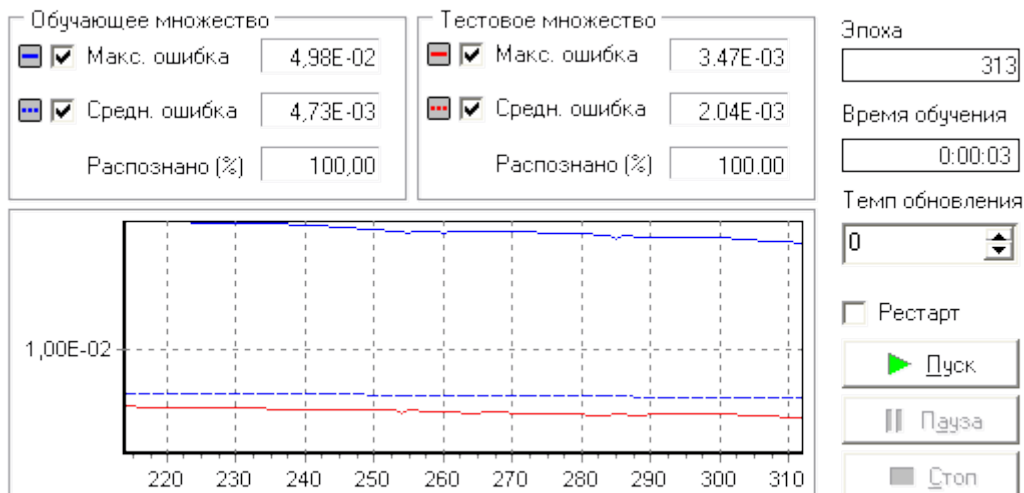


Рис.41

Обучение остановлено на 313 эпохе, т.е. максимальная ошибка стала меньше 0,05. Как видно из результатов обучения, были распознаны все элементы обучающего и тестового множеств, значит изменение параметров нейросети не требуется.

Далее к этой ветке сценария добавляем обработчик "Прогнозирование", где задаем горизонт прогноза, равный 4, т.е. строим прогноз на 4 последующих месяца (рис.42).

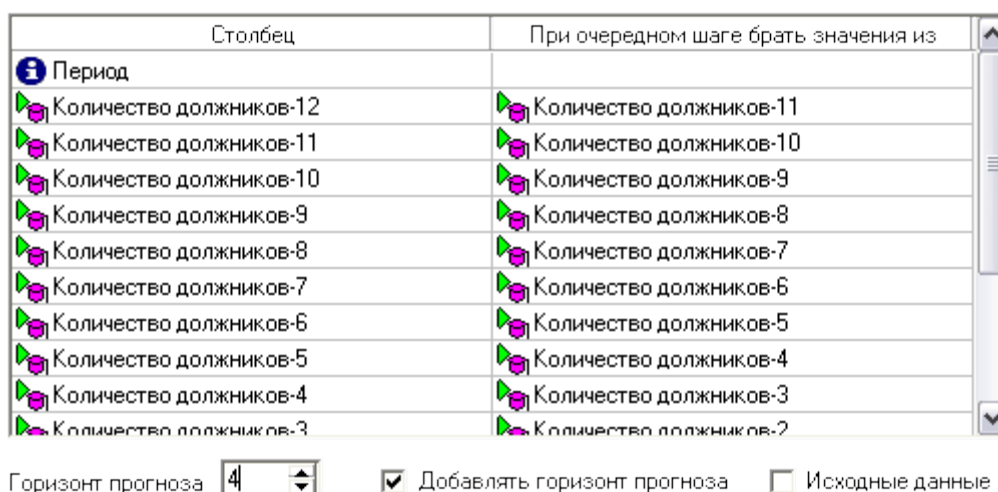


Рис.42

Получился сценарий (рис.43).

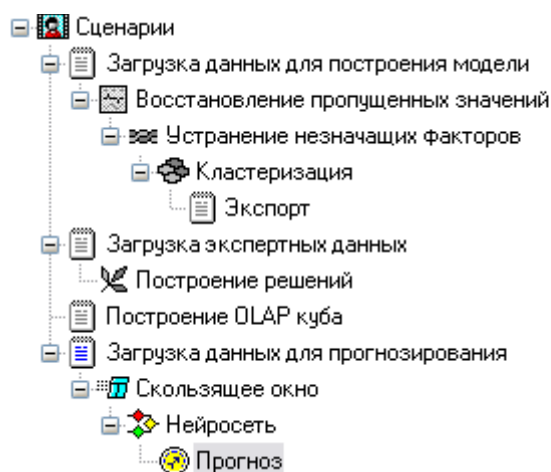


Рис.43

Выбираем визуализатор "Таблица" и оставляем в ней только поля "Количество должников" и "Шаг прогноза".

| Количество должников | Шаг прогноза |
|----------------------|--------------|
| 93                   | 1            |
| 94                   | 2            |
| 86                   | 3            |
| 82                   | 4            |

Таким образом нам удалось построить прогноз количества должников на апрель, май, июнь и июль 2021 года.

#### Задание.

Постройте прогноз на год, используя данные за 3 года, приведенные в файле "Task 6.txt", расположенного в папке примеров Deductor.

**Содержание отчета:** Отчёт по выполненной работе.

**Контрольные вопросы** (перечень вопросов по теме, на которые студент обязан знать ответы) и /или тестовые задания

1. Сформулируйте прикладную экономическую или управленческую оптимизационную задачу и опишите ее решение с применением генетического алгоритма.
2. Классифицирующие системы Холланда.
3. Перечислите основные этапы технологии генетического программирования.

## Список рекомендуемой литературы

### Основная литература:

1. Воронова Л.И., Воронов В.И. Big Data. Методы и средства анализа: Учебное пособие / МТУСИ. – М., 2016. – 35 с.
2. Суляев, И. И. Визуализация систем управления Электронный ресурс / Суляев И. И. : практикум. - Норильск : НГИИ, 2017. - 176 с. - ISBN 978-5-89009-686-9, экземпляров неограничено
3. Сузи,, Р. А. Язык программирования Python : учебное пособие / Р. А. Сузи. - Язык программирования Python, 2022-07-28. - Электрон. дан. (1 файл). - Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. - 350 с. - электронный. - Книга находится в премиум-версии ЭБС IPRBOOKS. - ISBN 978-5-4497-0705-5, экземпляров неограничено

### Дополнительная литература:

1. Тягунов, Г.В. Математические методы анализа: учебное пособие для СПО / Е.А. Трофимова, С.В. Плотников, Д.В. Гилёв; под. ред. Е.А. Трофимовой. —Эл. изд.— 2-е изд., стер.—Саратов : Профобразование; Екатеринбург : Изд-во Урал. ун-та, 2019. — 271с.
2. Суляев, И. И. Визуализация систем управления Электронный ресурс / Суляев И. И. : практикум. - Норильск : НГИИ, 2017. - 176 с. - ISBN 978-5-89009-686-9, экземпляров неограничено
3. Учебно-методическое пособие по дисциплине Интеллектуальные информационные системы и технологии Электронный ресурс / сост. Е. Н. Турута. - Учебно-методическое пособие по дисциплине Интеллектуальные информационные системы и технологии, 2022-04-04. - Москва : Московский технический университет связи и информатики, 2014. - 24 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 2227-8397, экземпляров неограничено

### Интернет-ресурсы:

1. Официальный сайт библиотеки ФГАОУ ВО СКФУ Режим доступа: <http://catalog.ncstu.ru/catalog> –.
2. Информационная справочная система ГАРАНТ.РУ // Режим доступа: <http://www.garant.ru/>
3. Информационная справочная система КонсультантПлюс. // Режим доступа: <http://www.consultant.ru>
4. Инфраструктура научно-исследовательских данных — платформа доступа к данным для научных исследований // Режим доступа: <https://www.data-in.ru>
5. Профессиональное сообщество специалистов по обработке данных и машинному обучению// Режим доступа: <https://www.kaggle.com/>
6. Профессиональная база данных Росстата // Режим доступа: Росстат — Базы данных ([rosstat.gov.ru](http://rosstat.gov.ru))