Министерство науки и высшего образования Российской Федерации

Изображение выглядит как текст, коллекция картинок

Автоматически созданное описание

Федеральное государственное бюджетное образовательное учреждение  
 высшего образования

Санкт-Петербургский горный университет  
императрицы Екатерины II

Кафедра информационных систем и вычислительной техники

**Расчетно-графическое задание (РГЗ)**

(Практические занятия № 13-17)

По дисциплине:

**ОСНОВЫ ТЕХНОЛОГИИ ПРОИЗВОДСТВА**

**ТЕМА**:

АНАЛИЗ СОВРЕМЕННОГО СОСТОЯНИЯ ПРОИЗВОДСТВЕННЫХ ПРОЦЕССОВ **(ПЗ №13)**

ПРИЕМЫ АВТОМАТИЗИРОВАННОГО УПРАВЛЕНИЯ ПРОИЗВОДСТВОМ **(ПЗ №14-17)**

|  |  |  |  |
| --- | --- | --- | --- |
| Выполнил: студент гр.\_ ИН-23-9\_\_\_\_\_\_\_\_ | | \_\_\_\_\_\_\_\_\_\_ | Миронов Д.В. |
| (шифр группы) | | (подпись) | (Ф.И.О) |
|  | |  |  |
| Проверил: | доцент | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | / А.В. Гурко / |
|  | (должность) | (подпись) | (Ф.И.О.) |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Санкт-Петербург

2023

# **Практическое занятие 13.** **Анализ современного состояния производственных процессов** **(Аналитическая платформа Deductor)**

Цели практической работы:

1. Познакомиться с архитектурой и пользовательским интерфейсом платформы Deductor, получить навыки создания сценариев обработки и визуализации данных.

Для получения ответов на вопросы получаем выполняем действия, указанные в вопросах. После этого анализируем полученный результат.

2.1. При импорте оставьте  **все настройки по умолчанию**. Ответьте на вопрос: *Какая проблема данных обнаружится?*

*Ответ: будет ошибка из-за невозможности преобразовать указанные в таблице “средняя продолжительность разговоров” данные в тип данных дата/время. Для решения этой проблемы необходимо изменить разделитель целой и дробной части с* «.» на «,».

4. Добавьте визуализатор *Статистика*. Просмотрите статистику по полю *Возраст*. *Каков минимальный и максимальный возраст абонентов?*

*Ответ: Максимальный возраст абонентов 70 лет, а минимальный 19.*

5. Просмотрите статистику по полю *Звонков днем за месяц*. Нажмите кнопку *Обзор статистики*на панели табличного визуализатора. *Сколько абонентов совершило 20 звонков днем за месяц?*

*Ответ: 20 звонков за день совершило 8 (0.2% от всех звонков днём за месяц) человек*

6. С помощью мастера обработки и метода  *Фильтрация*  отфильтруйте исходный набор данных по полю  *Возраст*  – выведите данные абонентов от 30 до 50 лет включительно.  *Сколько записей прошло через фильтр?*

*Ответ: 2231 записей*

7. С помощью мастера обработки и метода  *Калькулятор*  добавьте вычисляемое поле

*Всего звонков за месяц*  =  *Звонков днем за месяц* + *Звонков вечером за месяц* + *Звонков ночью за месяц*.  *Сколько всего звонков за месяц совершил абонент с кодом 10?*

*Ответ: 309 звонков*

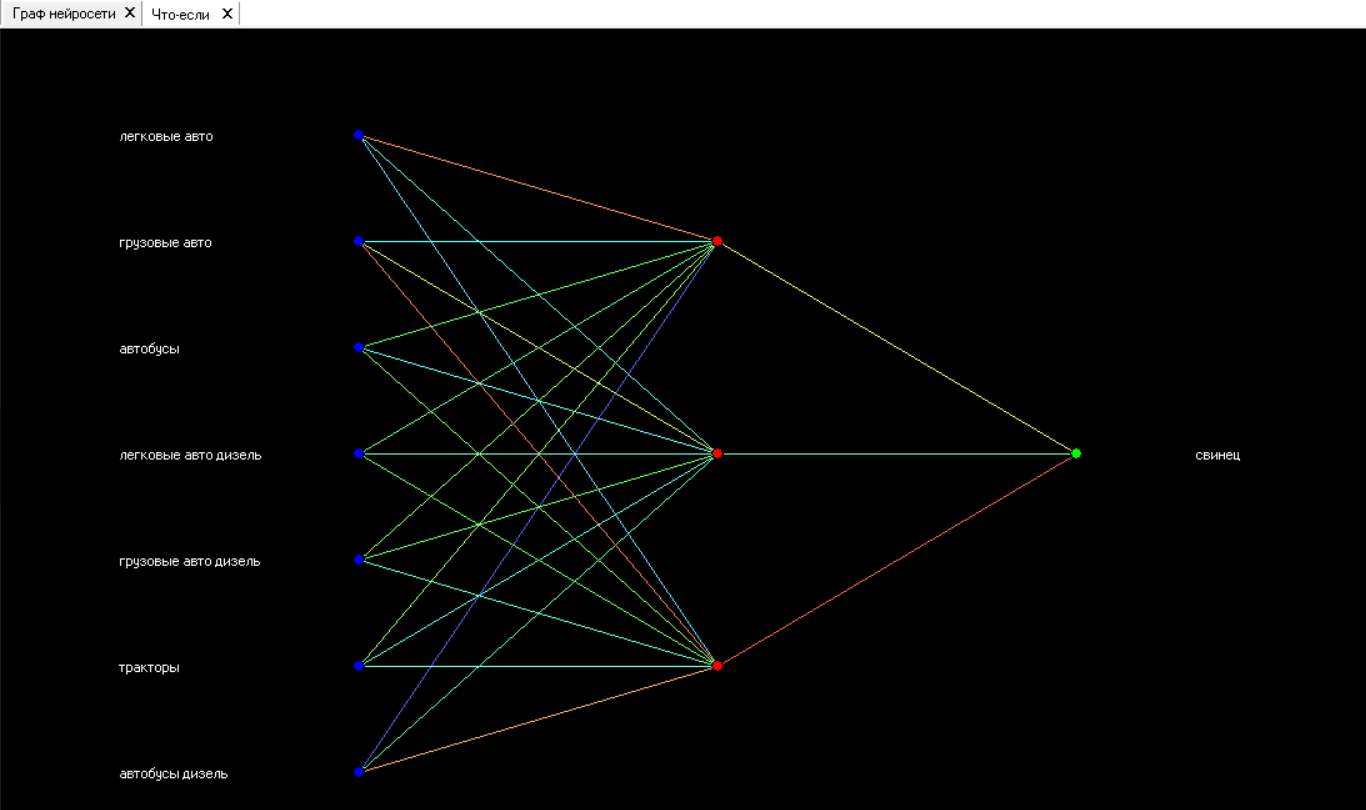
# **Практическое занятие 14.** **Методы анализа технологических процессов** **(Персептрон)**

Цели практической работы:

1. Познакомиться с архитектурой и пользовательским интерфейсом платформы Deductor, получить навыки создания сценариев обработки и визуализации данных.

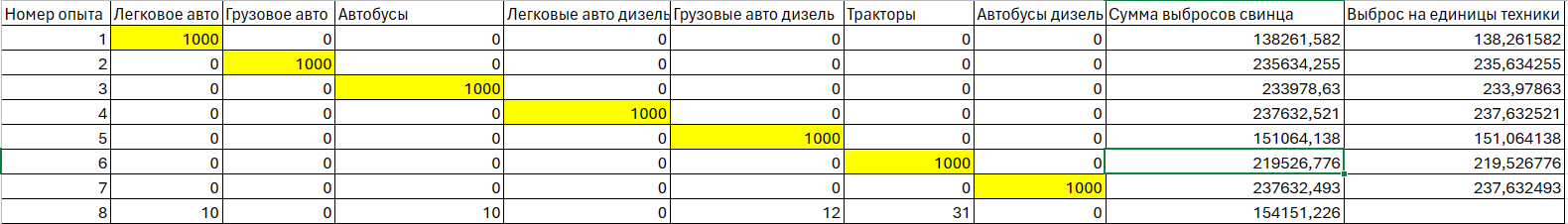
**Задача. Определить выброс свинца в атмосферу в зависимости от количества и вида проезжающего транспорта.**

После выполнения заданных в условии пунктов, я получил граф нейросети и визуализатор “что-если”, который может её настраивать.



Чтобы определить зависимость выброса свинца в атмосферу, нужно приравнять количество транспорта всех видов к нолю, кроме одного и сделать его количество равным одной тысячи (необходимо взять большое значение, чтобы получить ход вещей приближенный к реальному. Так мы получим качественное понимание того, что происходит). Таким образом мы и получим данные о загрязнении от одного вида транспорта.

Если приравнять весь транспорт к 0, то выброс свинца составит 42295,038. От этого значения мы будет отталкиваться при высчитывании степени влияния единицы транспорта на выброс свинца. Чтобы было легче вести записи и подсчёт, создадим таблицу в Excel’е.



Проанализировав полученную таблицу, я определил, что легковое авто выбрасывает в атмосферу меньше всего (138,261582), грузовое авто (235,634255), автобусы (233,97863), легковые авто дизель (237,632521), грузовые авто дизель (151,064138), тракторы (219,526776), автобусы дизель (237,632493) - из этих данных можно получить количество выбрасываемого свинца в зависимости от вида транспорта. Исходя из данных полученных в ходе опытов, я могу уверено заявить, что чем больше транспорта и чем больше его видов с высоким выбросом свинца, тем больше его выброс в атмосферу.

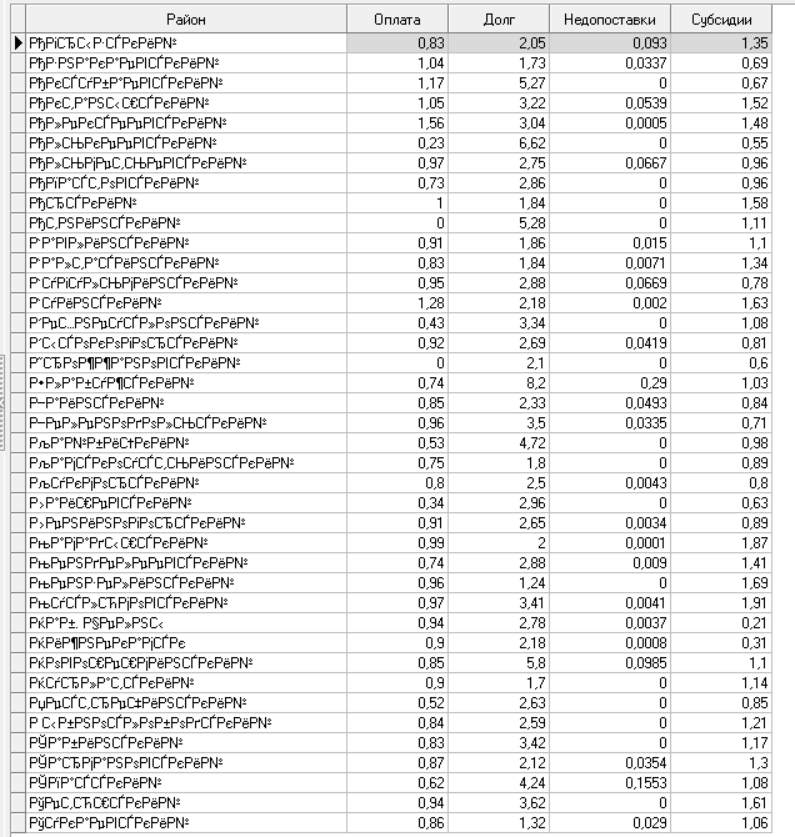
# Практическое занятие 15. Методы анализа технологических процессов (Кластеризация данных )

Цели практической работы:

1. Познакомиться с подходом к решению задачи кластеризации.
2. Приобрести навык самостоятельного решения задачи кластеризации

При открытии файла txt, полученного из исходного предоставленного файлы, слова превращались в непонятные символы, но это никак не влияет на результат, поэтому я изменил название только у главных столбцов в Deductor Studio Academic. Далее проблема будет устранена с помощью экспорта из офиса Excel в файл txt.

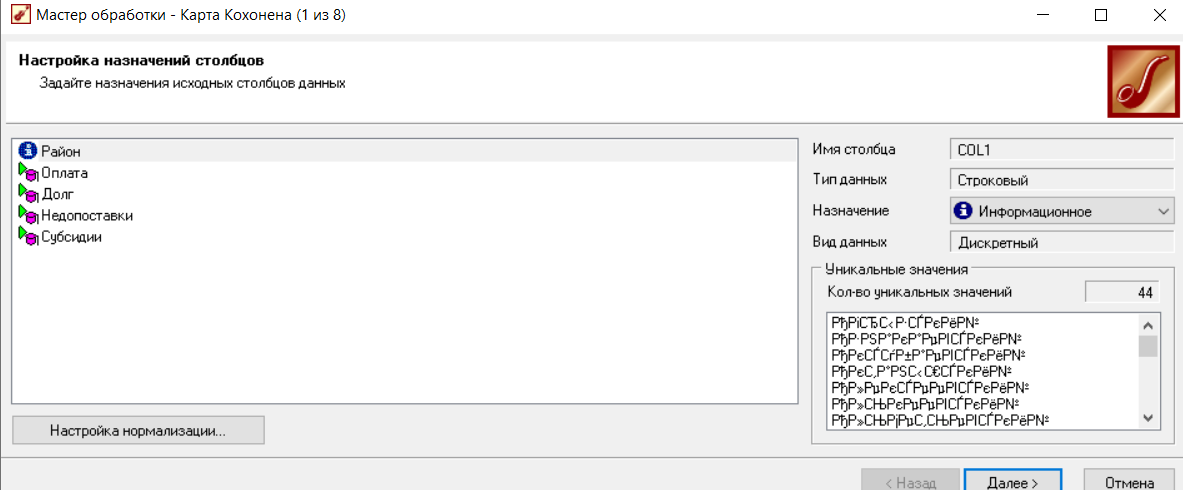
## **IV. Импорт данных**



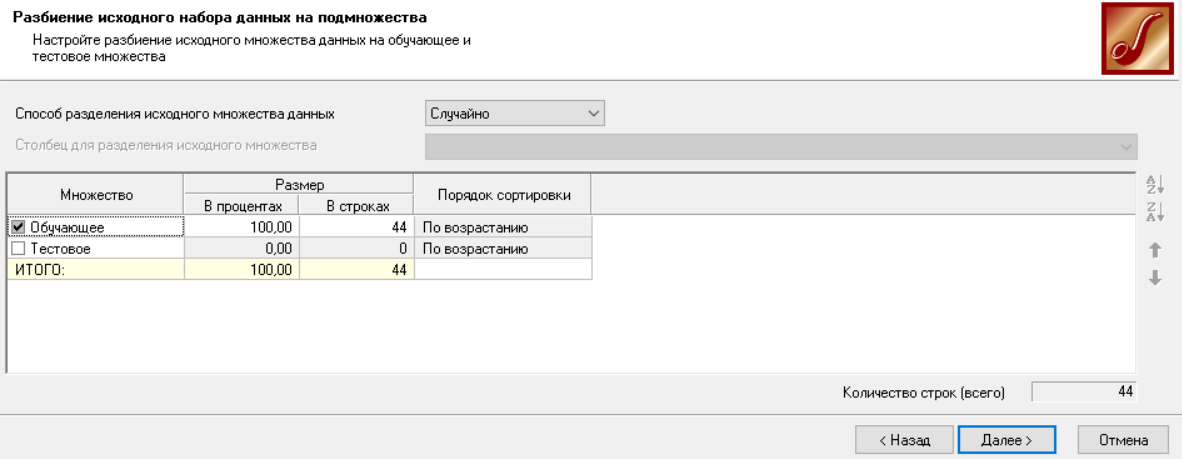
**V. Обработка данных**

Выполняя действия по указаниям, выполним настройку карты Кохонена.

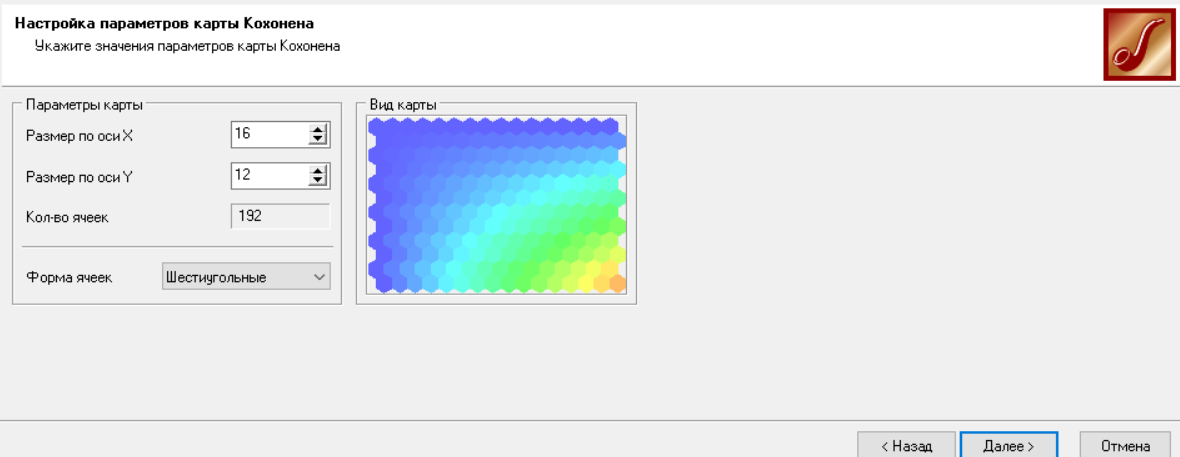
Шаг 1 из 8

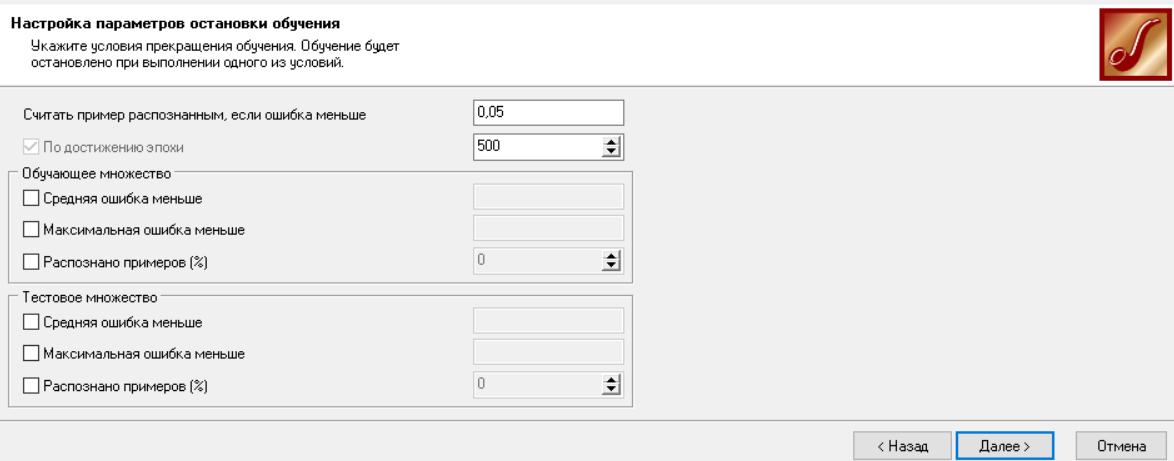


Шаг 2 из 8

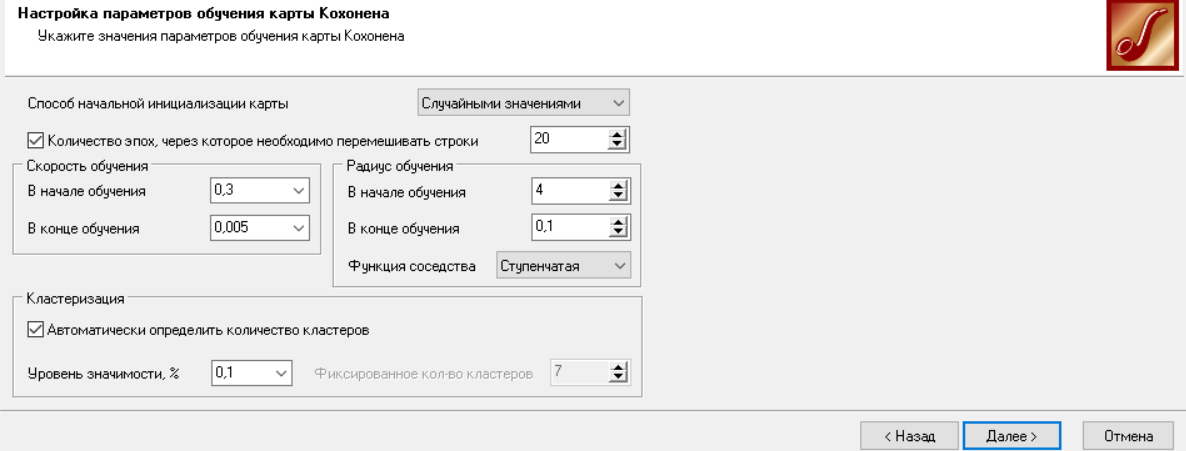


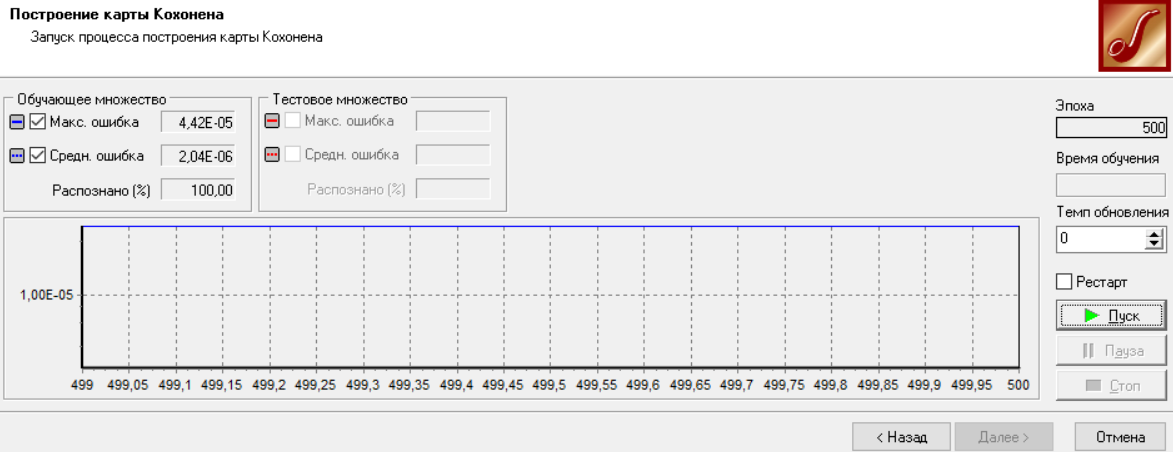
Шаг 3 из 8

Шаг 4 из 8

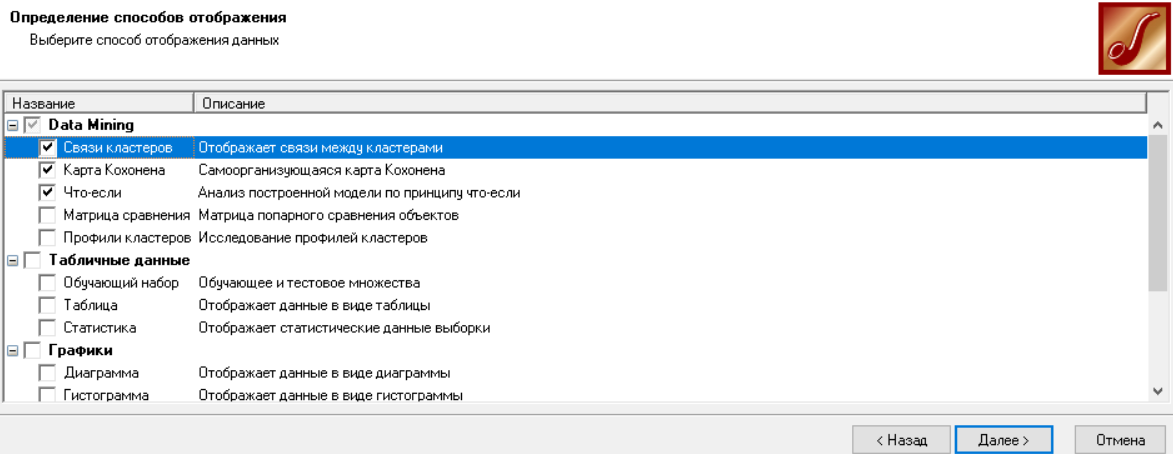


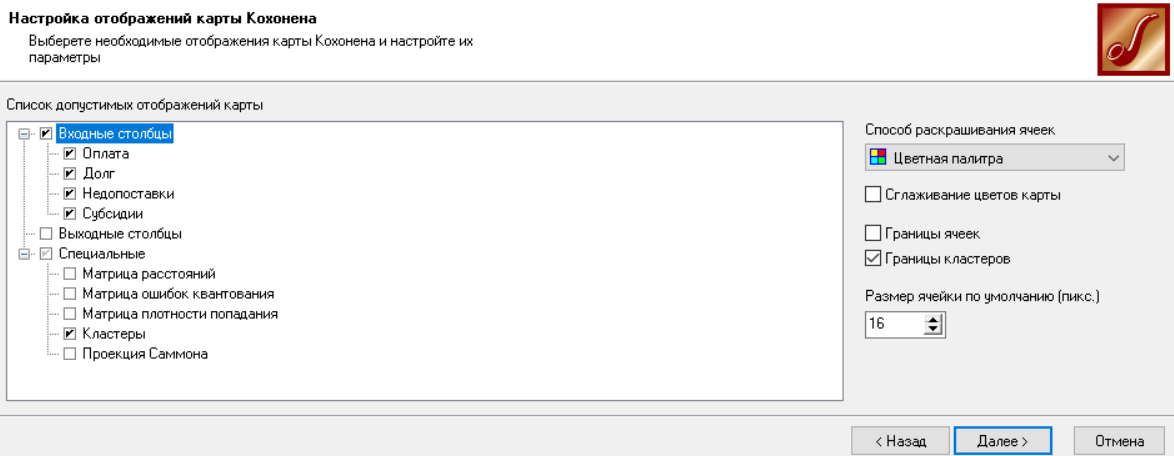
Шаг 5 из 8

Шаг 6 из 8



Шаг 7 из 8

Шаг 8 из 8



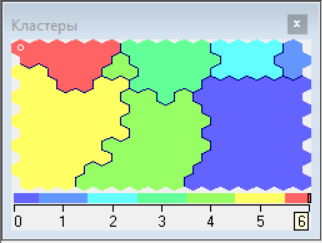
**VI. Анализ полученных данных**



Обучив сеть и настроив ее отображения, можно приступить к анализу и интерпретации результатов кластеризации. На рис. 16 показаны карты входов – оплата, долги, недопоставки, субсидии.

На одной из карт выделяем область с наибольшими значениями показателя. Далее имеет смысл изучить эти же нейроны на других картах.

На первой карте, отображающей оплату в районах, наибольшие значения имеют объекты, расположенные справа снизу. Рассматривая одновременно четыре карты, мы можем сказать, что эти же объекты имеют наибольшие значения показателя, изображенного на четвертой карте - субсидии. Долги же и недопоставки в районах этого кластера низкие. То есть, районы, попавшие в этот кластер, характеризуются как благоприятные. Это лишь фрагмент вывода, который можно сделать, исследуя карту.



Для нахождения конкретного объекта на карте необходимо нажать правой кнопкой мыши на исследуемом объекте и выбрать пункт "Найти ячейку на карте". Выполнение этой процедуры показано на рис. 18. В результате мы можем видеть как сам объект, так и значение того измерения, которое мы просматриваем. Таким образом, мы можем оценить положение анализируемого объекта, а также сравнить его с другими объектами.

В результате применения самоорганизующихся карт многомерное пространство входных факторов было представлено в двумерном виде, в котором его достаточно удобно анализировать.

# Практическое занятие 16. Методы анализа технологических процессов (Кластеризация данных )

Цели практической работы:

1. Познакомиться с подходом к решению задачи кластеризации.
2. Приобрести навык самостоятельного решения задачи кластеризации.

Рассматриваемые вопросы:

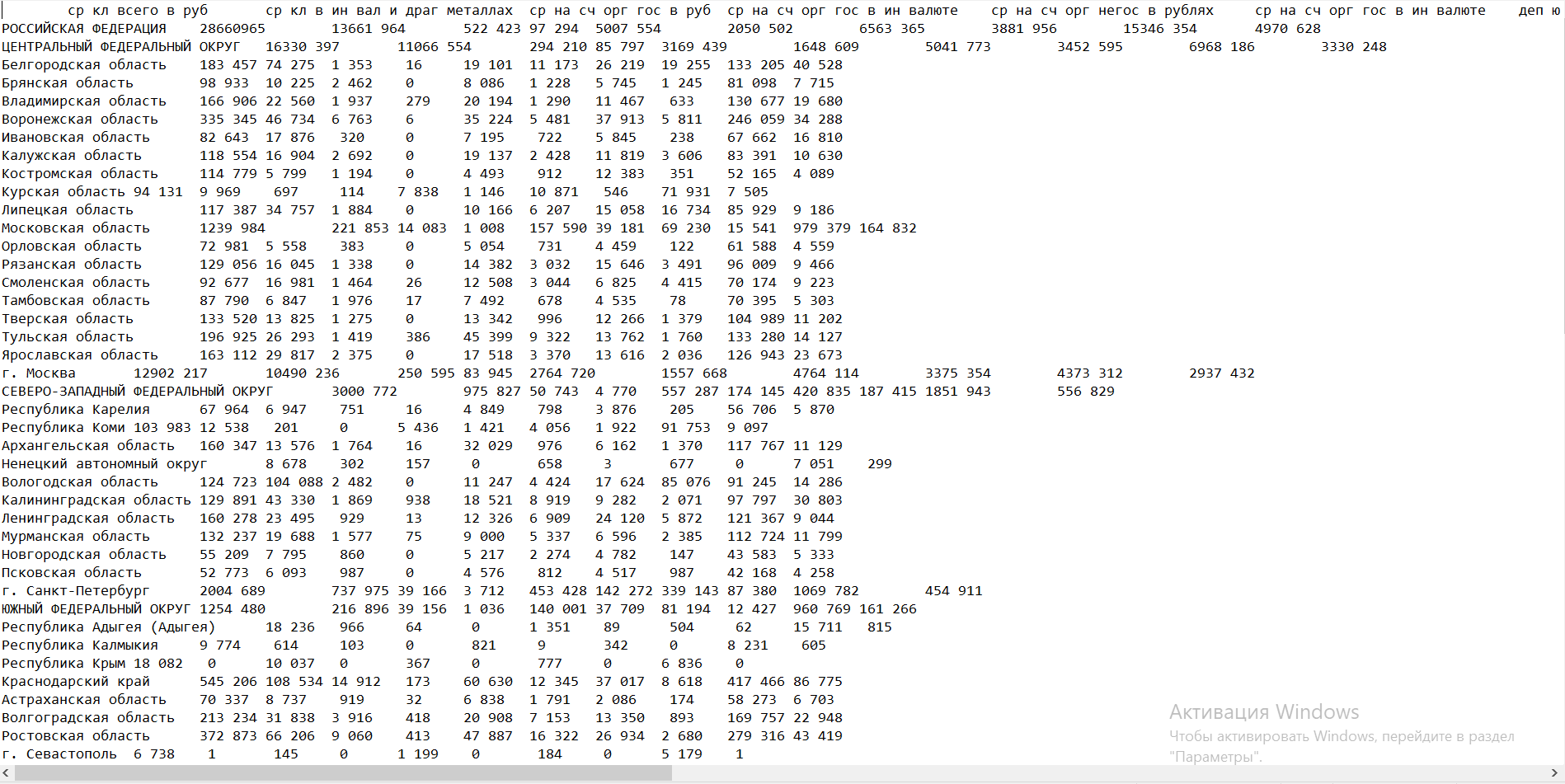
1. Решить задачу кластеризации
2. Ответить на вопросы
3. Подготовить отчет.

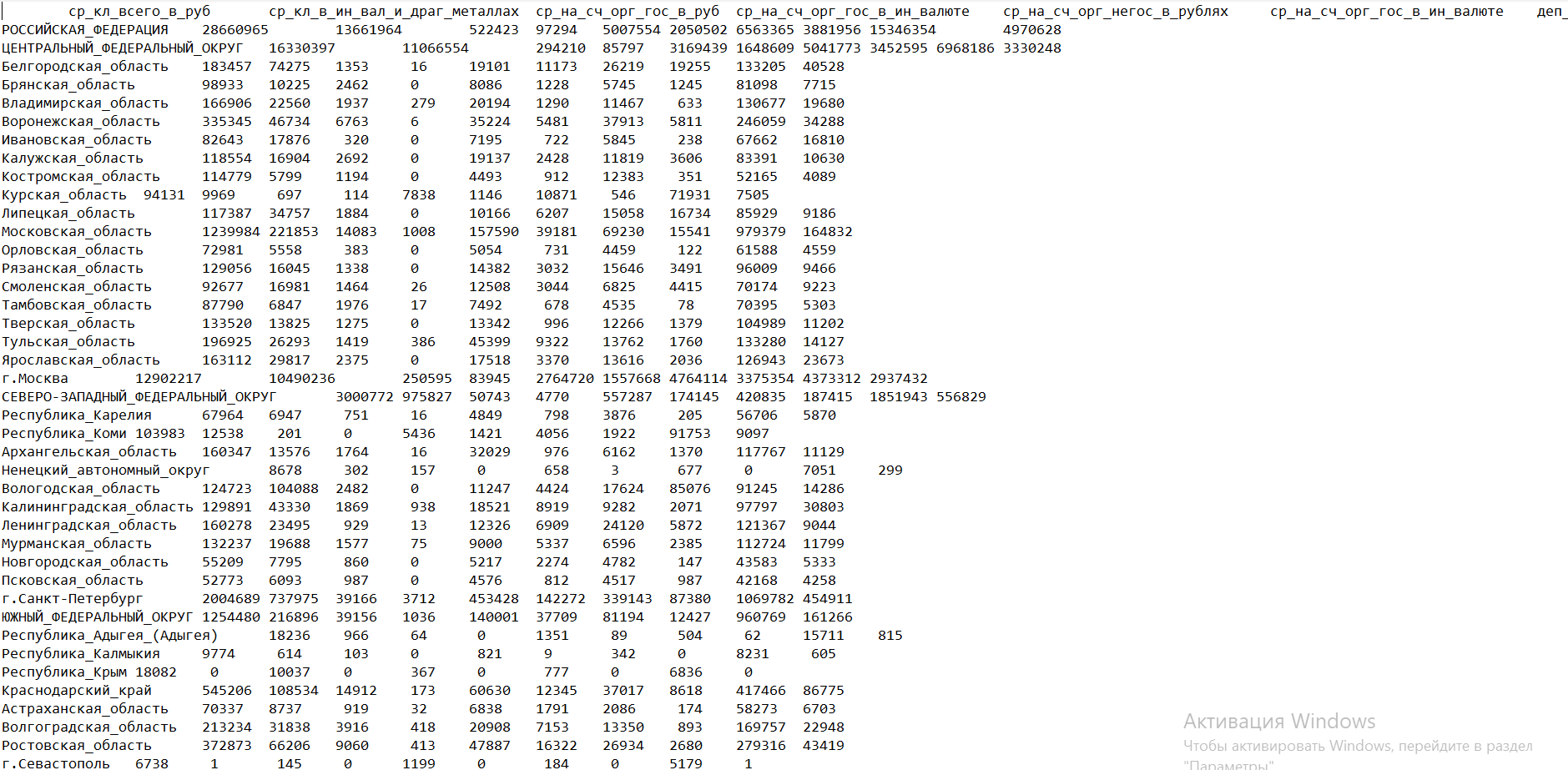
## Задание 4

### Исходные данные

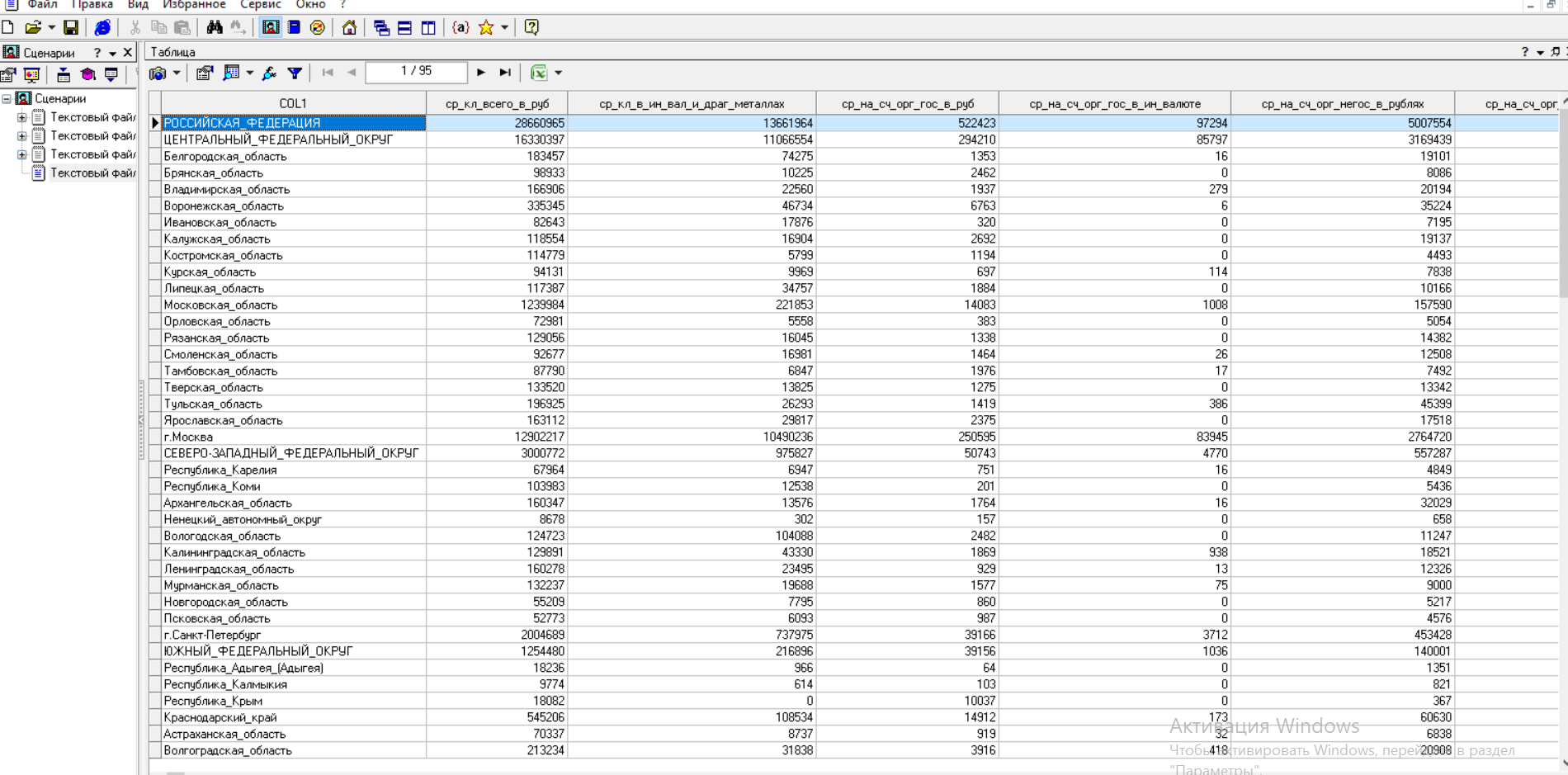
Спроектировать, обучить и сохранить нейронную сеть Кохонена для кластеризации субъектов РФ по объему средств в Deductor. Исходные данные содержатся в файле [ЛР8\_Банковские\_депозиты.xlsx](https://stepik.org/media/attachments/course/61480/%D0%9B%D0%A08_%D0%91%D0%B0%D0%BD%D0%BA%D0%BE%D0%B2%D1%81%D0%BA%D0%B8%D0%B5_%D0%B4%D0%B5%D0%BF%D0%BE%D0%B7%D0%B8%D1%82%D1%8B.xlsx):

Первым делом необходимо поработать над таблицей, приведя её в подобающий вид, а именно убрать все лишние строки в шапке таблицы, сделав так, чтобы над каждым столбцом данных располагалась только одна ячейка с данными о нём. Далее будет приведено фото того, что должно получиться после переделывания таблицы и конвертировании её в txt.

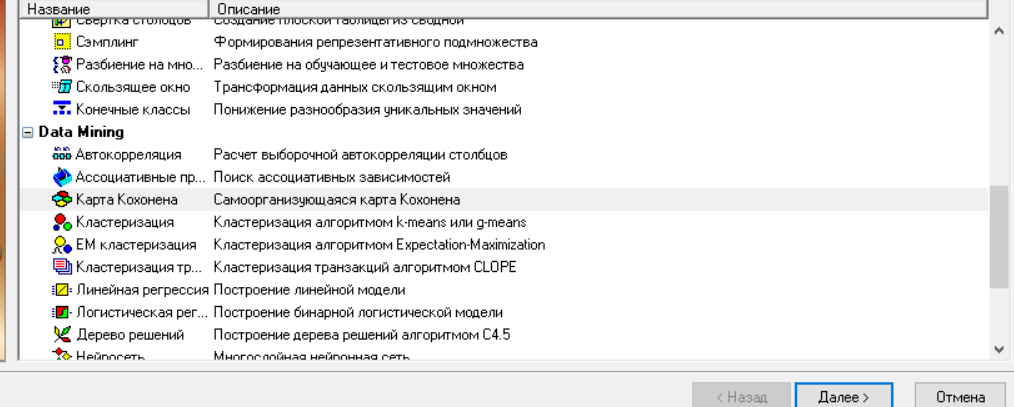


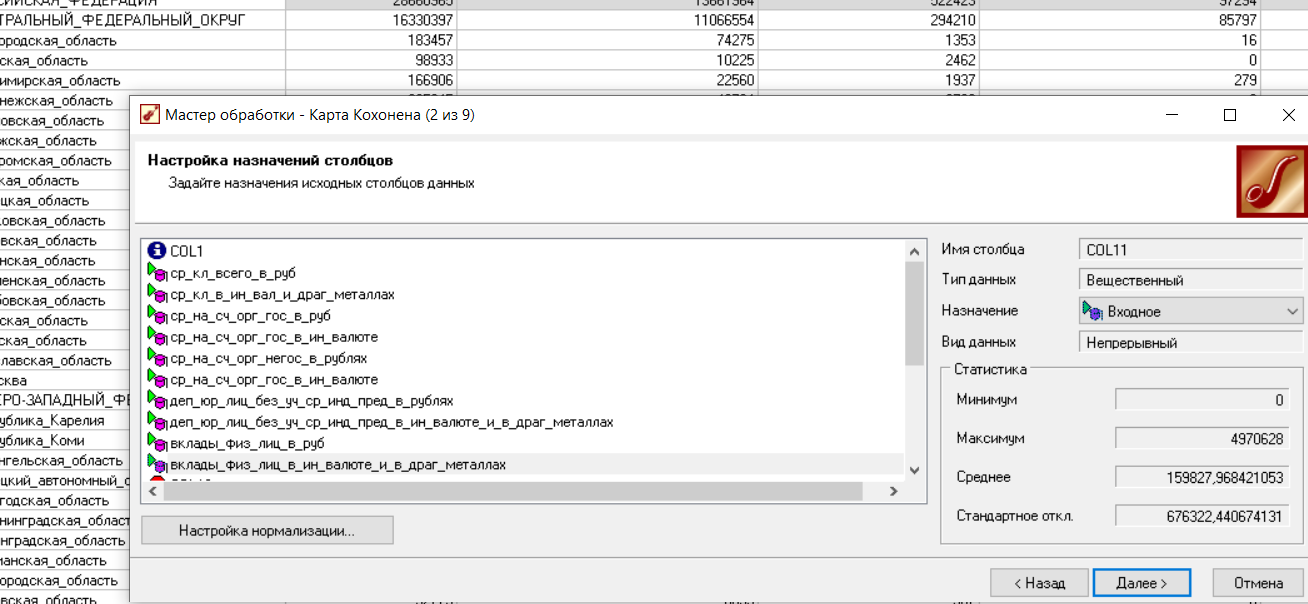
Далее необходимо импортировать ей в Deductor Studio Academic с помощью мастера импорта, но перед этим привести её в корректное состояние так, чтобы числа не разделились и так, чтобы у них не было ошибки при установлении типа данных “вещественные”.

После импортируем его в Deductor Studio Academic.



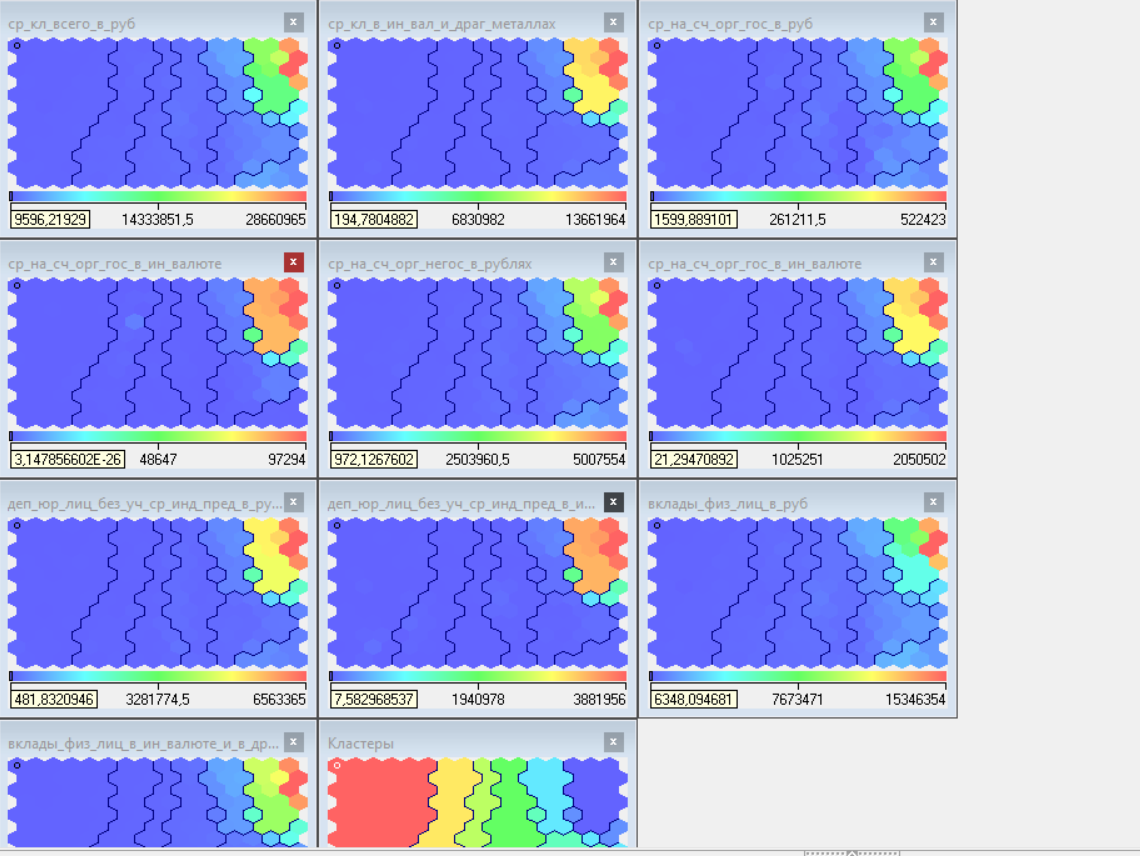
Далее используем мастер обработки, в нём в группе Data Mining выбираем Карта Кохонена и настраиваем.





Остальные параметры настраиваем по инструкции, данной в задании.

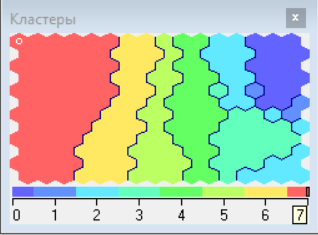
Получаем карту Кохонена.



После проведенной кластеризации выведите карты входов и проанализируйте полученные результаты.

### Контрольные вопросы:

1. Сколько получилось кластеров?

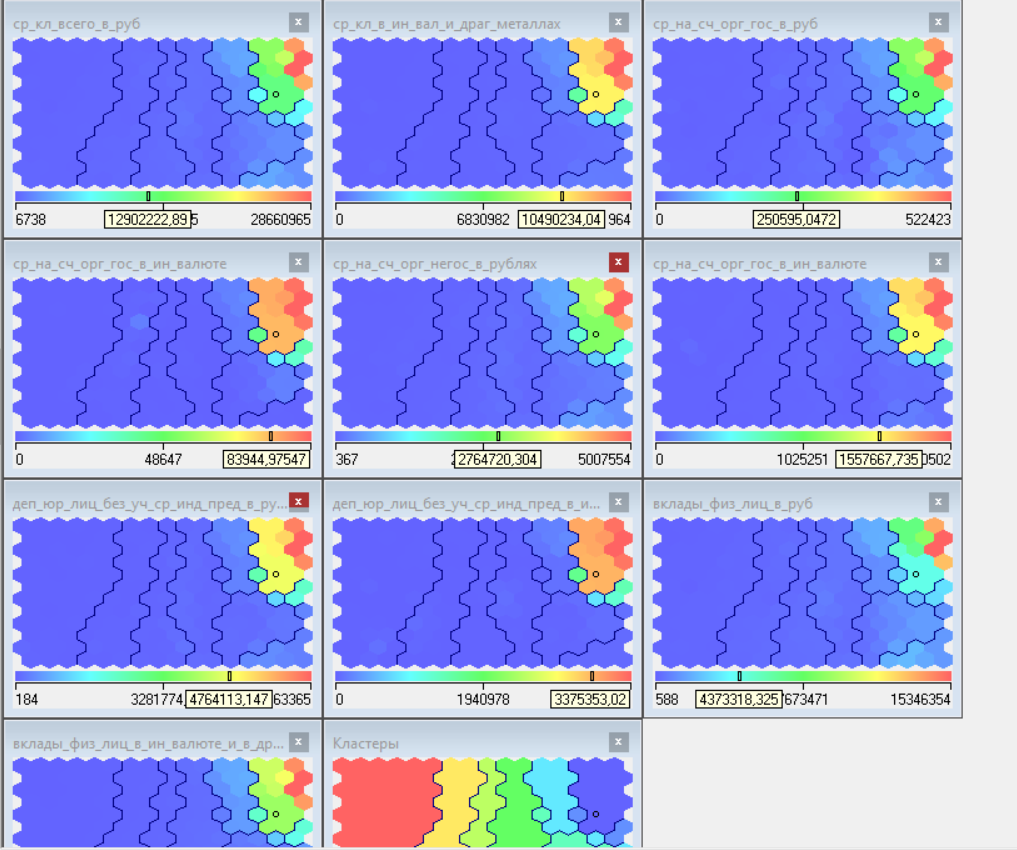


Ответ: всего кластеров получилось 8

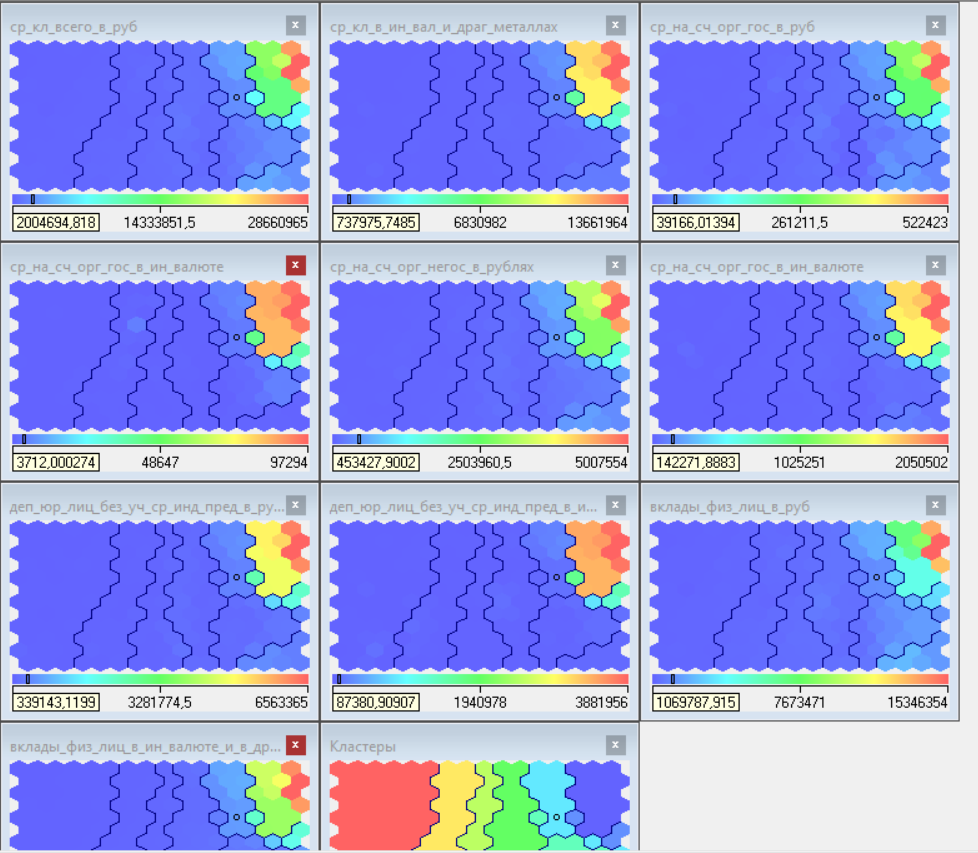
1. Найдите положение на карте г. Москва, г. Санкт-Петербург, Республики  Татарстан и г. Севастополь.

Ответ:

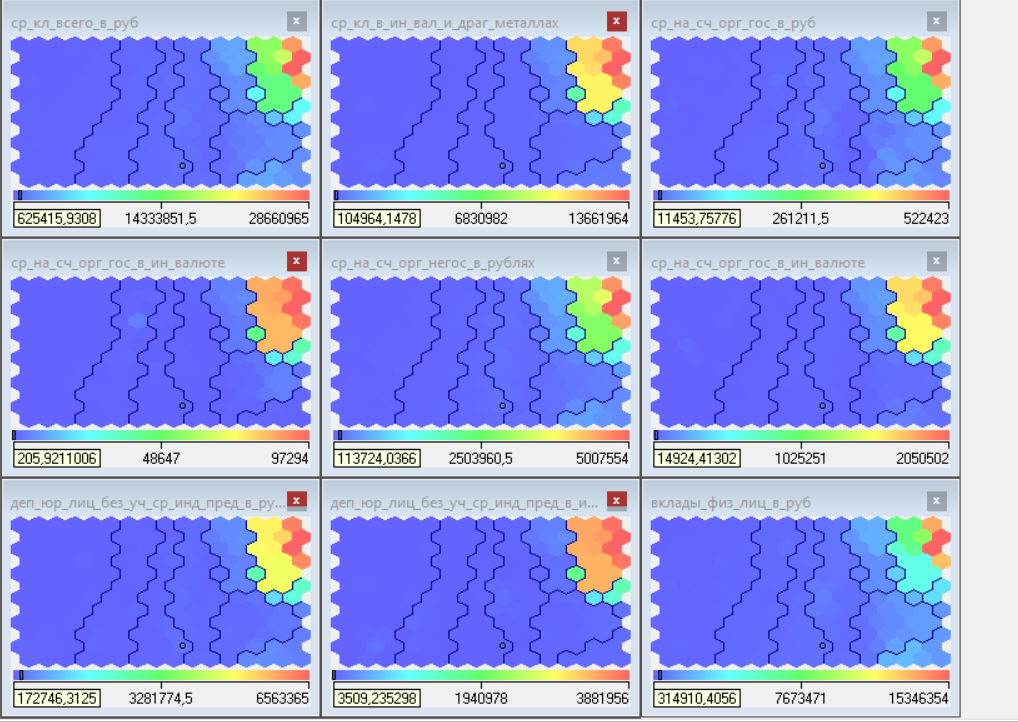
г.Москва:



Г.СПБ



Республика  Татарстан



г. Севастополь



1. Они попали в разные кластеры или в один?

Ответ: г.Москва попал в разные кластеры. Г. СПБ попал в один кластер. Республика Татарстан и г. Севастопль тоже попали в один кластер.

1. Как Вы думаете, почему?

Ответ: Каждый кластер имеет свой цвет и эти цвета обозначают определённые диапазоны значений. Например, есть два кластера: зелёный и голубой. Зелёный начинается от 0 и заканчивается в 3, а голубой начинается в 3 и заканчивается в 6, что говорит о том, что если у города будет 2, то он будет в зелёном кластере, а если 5, то в голубом. Аналогичная ситуация происходит с нашими объектами. Все представленные данные о регионах говорят о их состоятельность, а значит логично предположить, что все данные имеют связь. Если среднее количество всего в рублях низкое, то и остальные показатели не будут высокими. Это всё даёт возможность говорить о конкретном месте города в классификации состоятельности или же назвать его показатели одним словом, отражающим реальный ход вещей, “средние”, “высокие”, “низкие” и т.д. Резюмируя всё вышесказанное, можно отметить, что, если какой-то объект со связанными данными попал в кластер по одному параметру, то с высокой вероятностью в этот же кластер попадут и другие, но с одной оговоркой. Такое возможно только в том случае, когда данные разных объектов сильно разнятся или же есть сравнительно остальных большие значения, сравнительно остальных средние и т.д. , что в нашем случае достигается. Ярким примером, когда это не происходит, является г. Москва. Все данные этого объекта сравнительно остальных большие, но, несмотря на это, их нельзя назвать одним словом, потому что показатели находятся на разных уровнях относительно других. Так у города Москва среднее количество в рублях можно назвать средним показателем, а среднее количество в иностранной валюте и драгоценных металлах выше среднего.

1. Какие регионы попали в один кластер с Республикой Татарстан?

Ответ: Белгородская\_область, Белгородская\_область, Белгородская\_область, Пермский\_край, Нижегородская\_область, Челябинская\_область, Новосибирская\_область.

# Практическое занятие 17. Методы анализа технологических процессов (Аналитическая платформа Deductor)

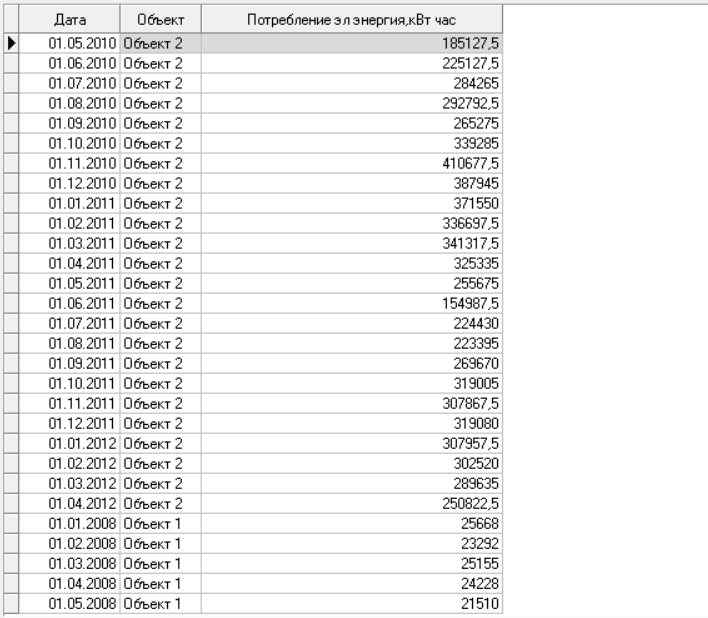
Цели практической работы:

1. Познакомиться с архитектурой и пользовательским интерфейсом платформы Deductor, получить навыки создания сценариев обработки и визуализации данных.

Рассматриваемые вопросы:

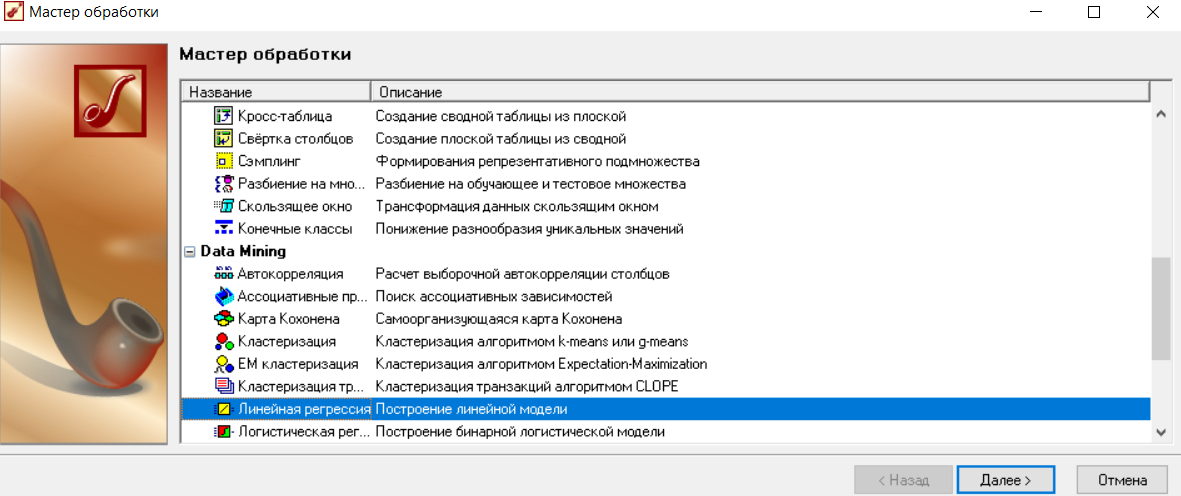
1. Линейная регрессия
2. Деревья решений

Сначала используем мастер импорта для того, чтобы импортировать файл с таблицей формата txt.



Предполагаем, что зависимость между входными факторами и результатом линейная, что значит линейная регрессия необходима.

Для построения линейной регрессии необходимо запустить Мастер обработки и выбрать в качестве обработки данных Линейную регрессию.



На первом шаге задаем назначение исходных столбцов. На следующем шаге происходит настройка обучающего и тестового множеств, способ разложения исходного множества данных. Третий шаг установки позволяет осуществить ограничение диапазона входных значений. Данный шаг оставим без изменений. При нажатии на кнопку "Далее" появляется окно запуска процесса обучения. В процессе выполнения видно, какая часть распознана на этапе обучения и теста. После выполнения процесса выберем в качестве способа отображения диаграмму рассеяния и отображение результатов в виде диаграммы. Как видно из диаграммы рассеяния, обучение прошло с хорошей точностью.

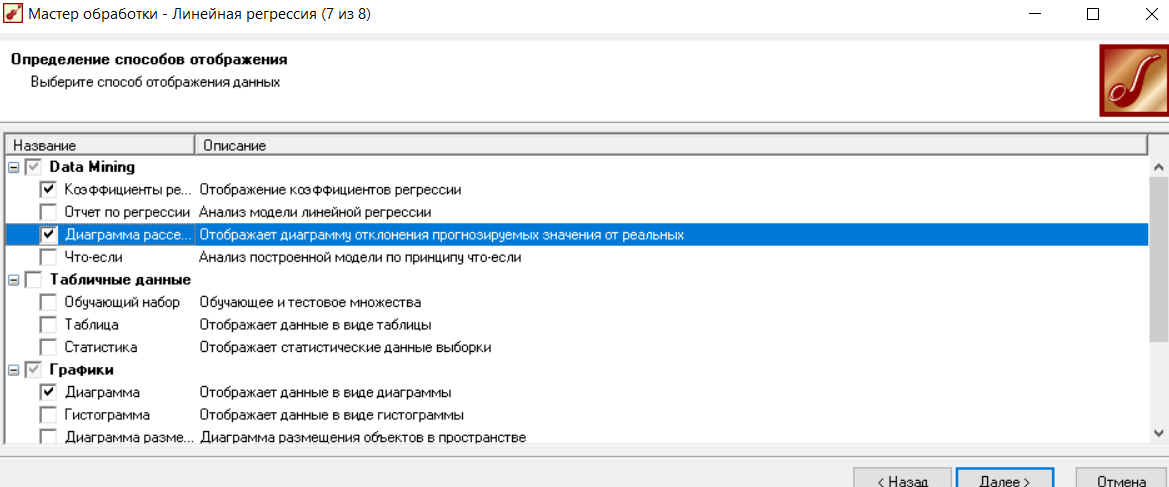


Диаграмма рессеяния:

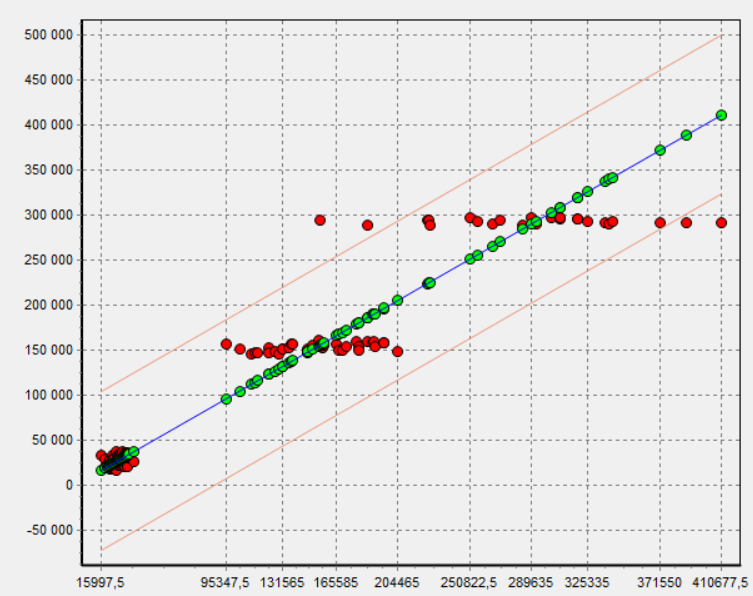
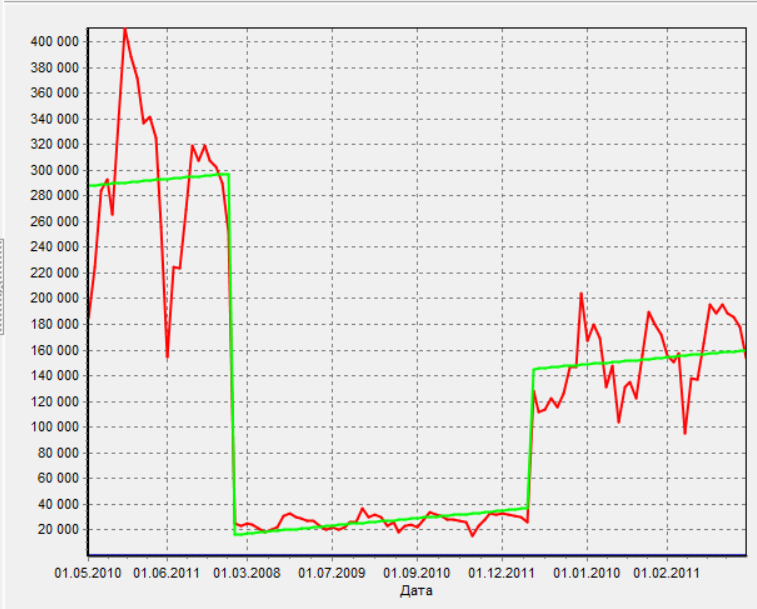
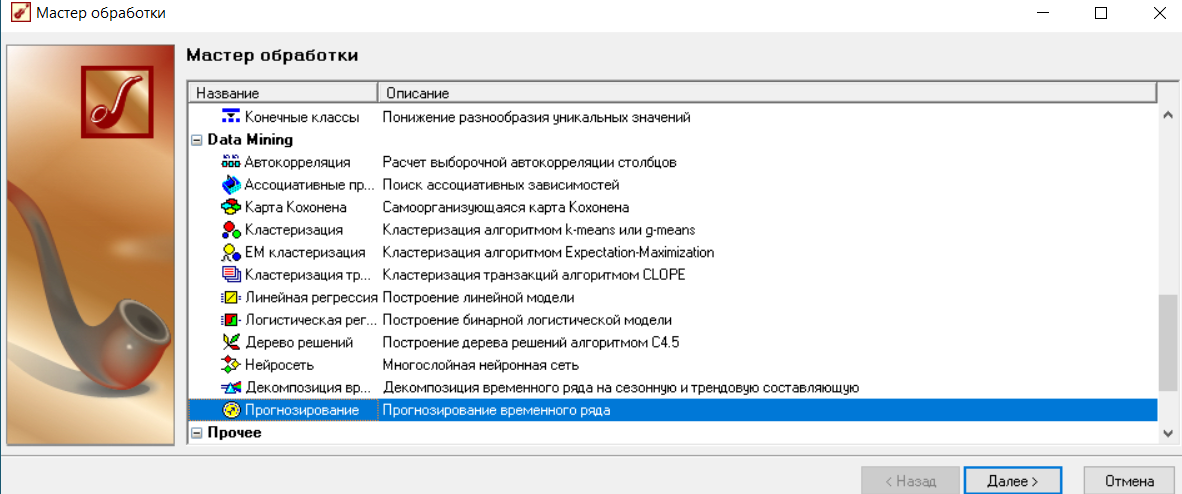


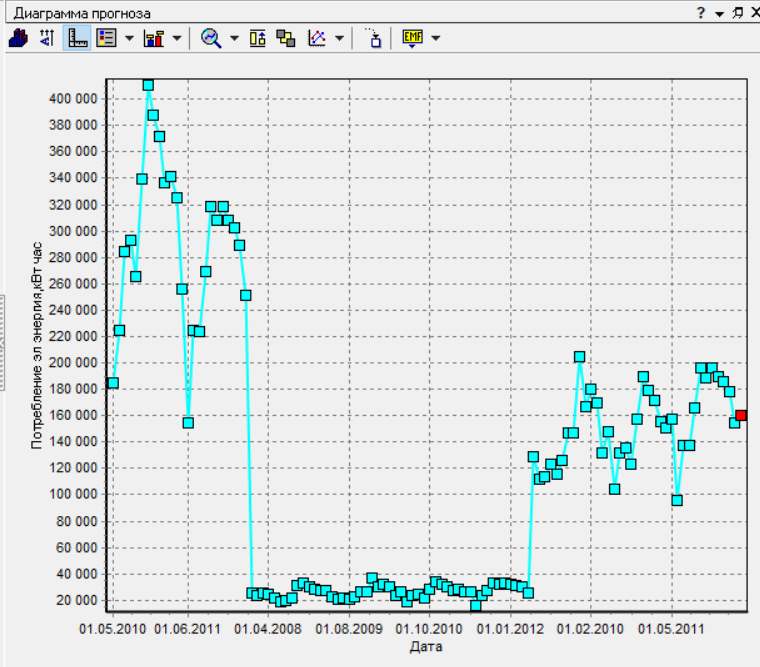
Диаграмма:



Прогнозирование. Теперь для построения прогноза запустим Мастера обработки, в котором выберем прогнозирование. На первом шаге обработчика происходит настройка связи столбцов для прогнозирования.



На следующем шаге задаются параметры визуализации. Для данного примера выбираем отображение результатов в виде диаграммы прогноза. Теперь аналитик может дать прогноз о продажах, основываясь на модели, построенной с помощью линейной регрессией.

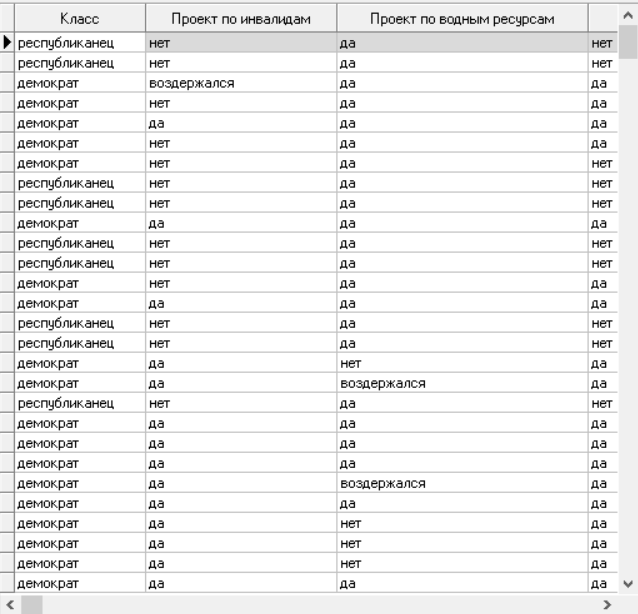
  
Выводы. Данный пример показал целесообразность применения линейного регрессионного анализа для прогнозирования линейных зависимостей.

Простота настроек и быстрота построения модели иногда бывают необходимы. Аналитику достаточно указать входные столбцы – факторы, выходные — результат, указать способ разбиения данных на тестовое и обучающее множество и запустить процесс обучения. Причем после этого будут доступны все механизмы визуализации и анализа данных, позволяющие построить прогноз, провести эксперимент по принципу "Что–если", исследовать зависимость результата от значений входных факторов, оценить качество построенной модели по диаграмме рассеяния. Также по результатам работы этого алгоритма можно подтвердить или опровергнуть гипотезу о наличии линейной зависимости.

# Деревья решений (Deductor)

Деревья решений применяются для решения задачи классификации. Дерево представляет собой иерархический набор правил, согласно которым данные относятся к тому или иному классу. В построенном дереве присутствует информация о достоверности того или иного правила. Рассчитывается значимость каждого входного поля.

Используем мастер импорта, чтобы импортировать таблицу 17P Потребление электрической энергии.txt.



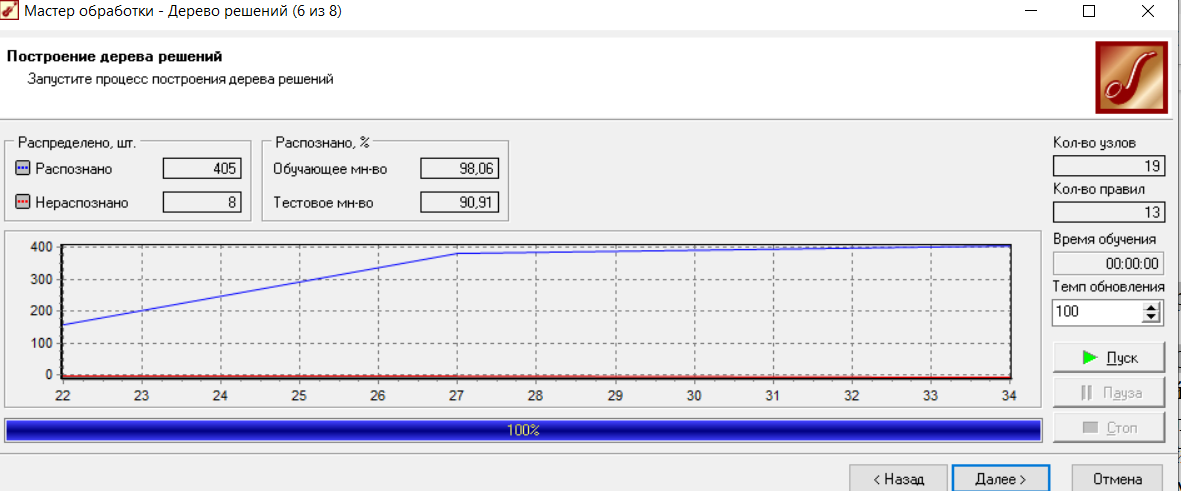
Для решения задачи запустим Мастер обработки. Выберем в качестве метода решения - дерево решений.

В Мастере построения дерева решения на втором шаге настроим поле "Код" информационным, "Класс" выходным, остальные поля входными.

Настроим способ разбиения исходного множества данных на обучающее и тестовое. Зададим случайный способ разбиения, когда данные для тестового и обучающего множества берутся из исходного набора случайным образом.

Настроим 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. Включим данные опции.

Выбираем процесс построения дерева решения в автоматическом режиме или интерактивном (полуавтоматическом). Выберем сначала автоматический режим построения и запустим. Далее можно увидеть информацию о количестве распознанных примеров.



После построения дерева можно увидеть, что почти все примеры и на обучающей и на тестовой выборке распознаны.

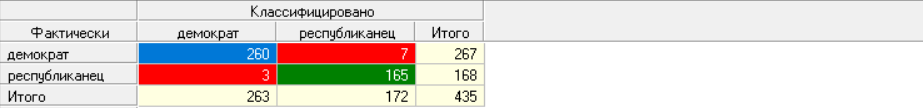
Выберем способ визуализации полученных результатов. Основной целью аналитика является отнесение депутата к той или иной партии. Механизм отнесения должен быть таким, чтобы депутат указал, как он будет голосовать за различные законопроекты, а дерево решений ответит на вопрос, кто он — демократ или республиканец. Такой механизм предлагает визуализатор "Что–если".

Выберем визуализатор "Дерево решений" и изучим дерево решений, на котором можно определить, какие факторы являются более важными (верхние узлы дерева), какие второстепенными, а какие вообще не оказывают влияния (входные факторы, вообще не присутствующие в дереве решений).

Формализованные правила классификации, выраженные в форме "Если <Условие>, тогда <Класс>", можно увидеть, выбрав визуализатор "Правила (дерево решений)". Часто аналитику полезно узнать, сколько примеров было распознано неверно, какие именно примеры были отнесены к какому классу ошибочно. На этот вопрос дает ответ визуализатор "Таблица сопряженности".

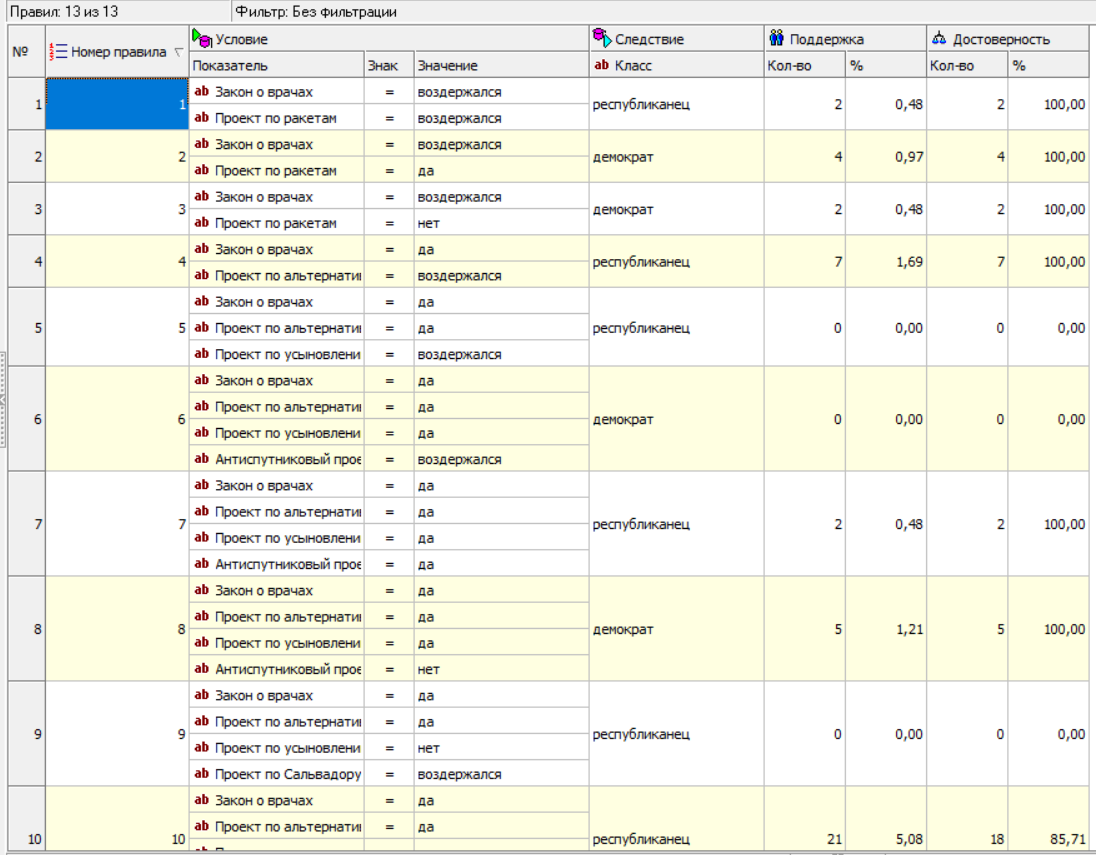
Важно знать, каким образом каждый фактор влияет на классификацию. Такую информацию предоставляет визуализатор "Значимость атрибутов".

Проанализируем данные при помощи имеющихся визуализаторов. Посмотрим на таблицу сопряженности.



По диагонали таблицы расположены примеры, которые были правильно распознаны, в остальных ячейках — те, которые были отнесены к другому классу. В данном случае дерево правильно классифицировало практически все примеры.

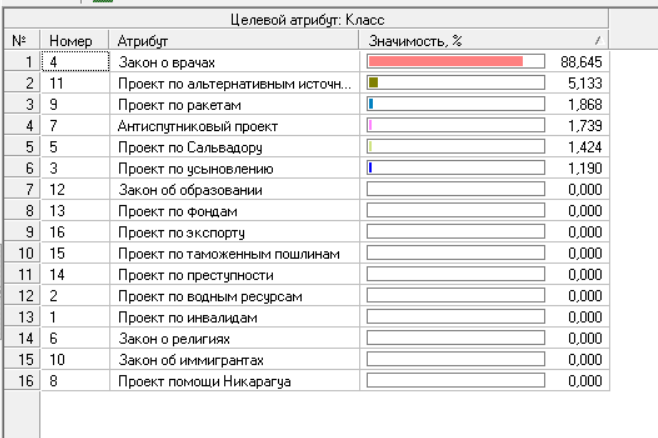
Перейдем к основному визуализатору — "Дерево решений". Дерево решений получилось простое, большая часть факторов (законопроектов) была отсечена, т. е. влияние их на принадлежность к партии минимальна или его вообще нет (по- видимому, по этим вопросам у партий нет принципиального противостояния).



Самым значимым фактором оказалась позиция, занимаемая депутатами по пакету законов, касающихся врачей, т. е. если депутат голосует против законопроекта о врачах, то он демократ

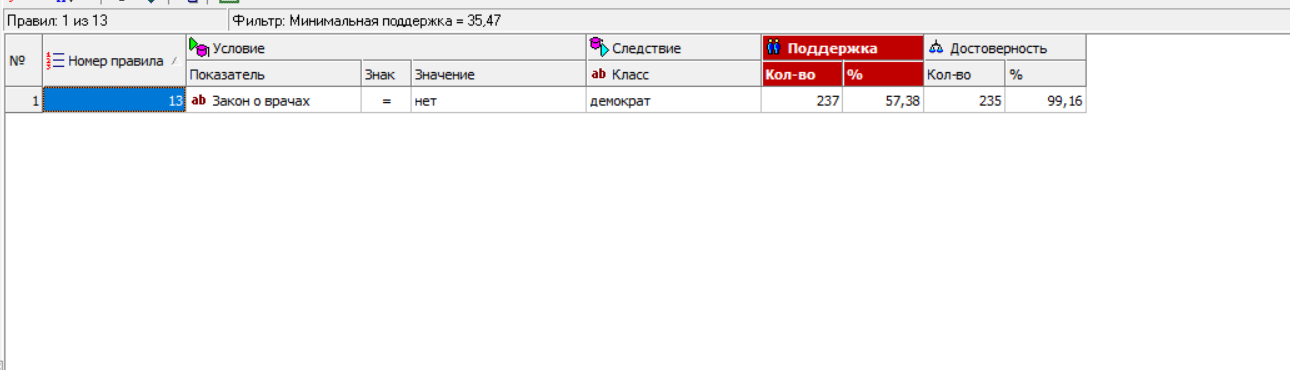
Достоверно судить о том, что депутат — республиканец, можно, если он голосовал за законопроект о врачах, а также за законопроект по Сальвадору, а также был против законопроекта об усыновлении. Данный визуализатор предоставляет возможность просмотра примеров, которые попали в тот или иной узел, а также информацию об узле.

Более удобно посмотреть значимость факторов или атрибутов в визуализаторе "Значимость атрибутов".



С помощью данного визуализатора можно определить, насколько сильно выходное поле зависит от каждого из входных факторов. Чем больше значимость атрибута, тем больший вклад он вносит при классификации. В данном случае самый большой вклад вносит закон о врачах, как и было сказано выше.

На визуализаторе "Правила" представлен список всех правил, согласно которым можно отнести депутата к той или иной партии. Правила можно сортировать по поддержке, достоверности, фильтровать по выходному классу (к примеру, показать только те правила, согласно которым депутат является демократом с сортировкой по поддержке).



Данные представлены в виде таблицы. Полями этой таблицы являются:

номер правила;

условие, которое однозначно определяет принадлежность к партии;

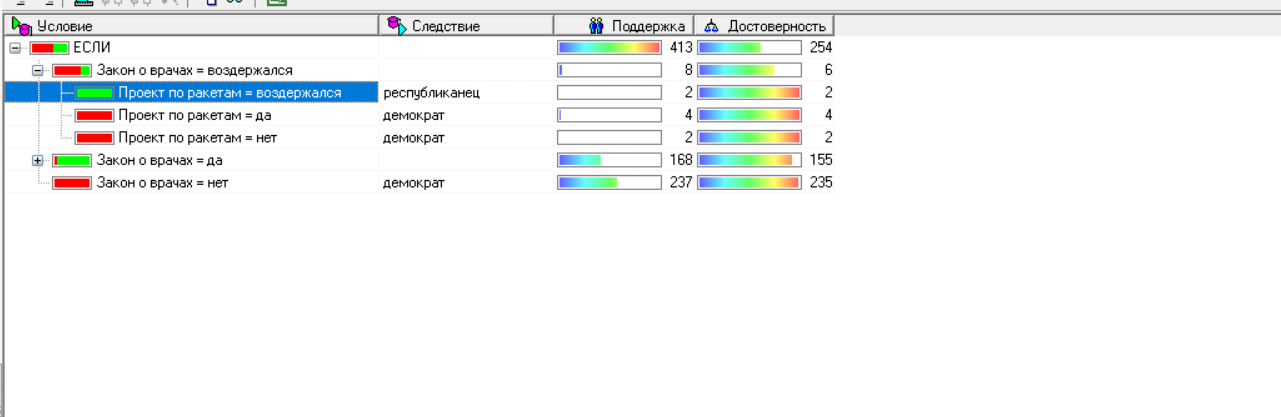
следствие — то, кем является депутат, голосовавший согласно этому условию;

поддержка — количество и процент примеров из исходной выборки, которые отвечают этому условию;

достоверность — процентное отношение количества верно распознанных примеров, отвечающих данному условию, к общему количеству примеров, отвечающих данному условию.

Исходя из данных этой таблицы, аналитик может сказать, что именно влияет на то, что депутат является демократом или республиканцем, какова цена этого влияния (поддержка) и какова достоверность правила. В данном случае совершенно очевидно, что из всего списка правил с достаточно большим доверием можно отнестись к двум: правилу №9 и правилу №7. Таким образом, получается, что демократы принципиально против законопроектов, касающихся врачей. Республиканцы же, наоборот, за принятие этих законопроектов и также за принятие законопроекта по Сальвадору, но категорически против законопроектов по усыновлению.

Теперь аналитик может точно сказать, кто есть кто. Но иногда аналитик считает правильным построить дерево решений исходя из своих соображений или внести некоторую корректировку, и тогда необходимо выбрать интерактивный режим построения, в результате чего получим следующее окно дерева решений.



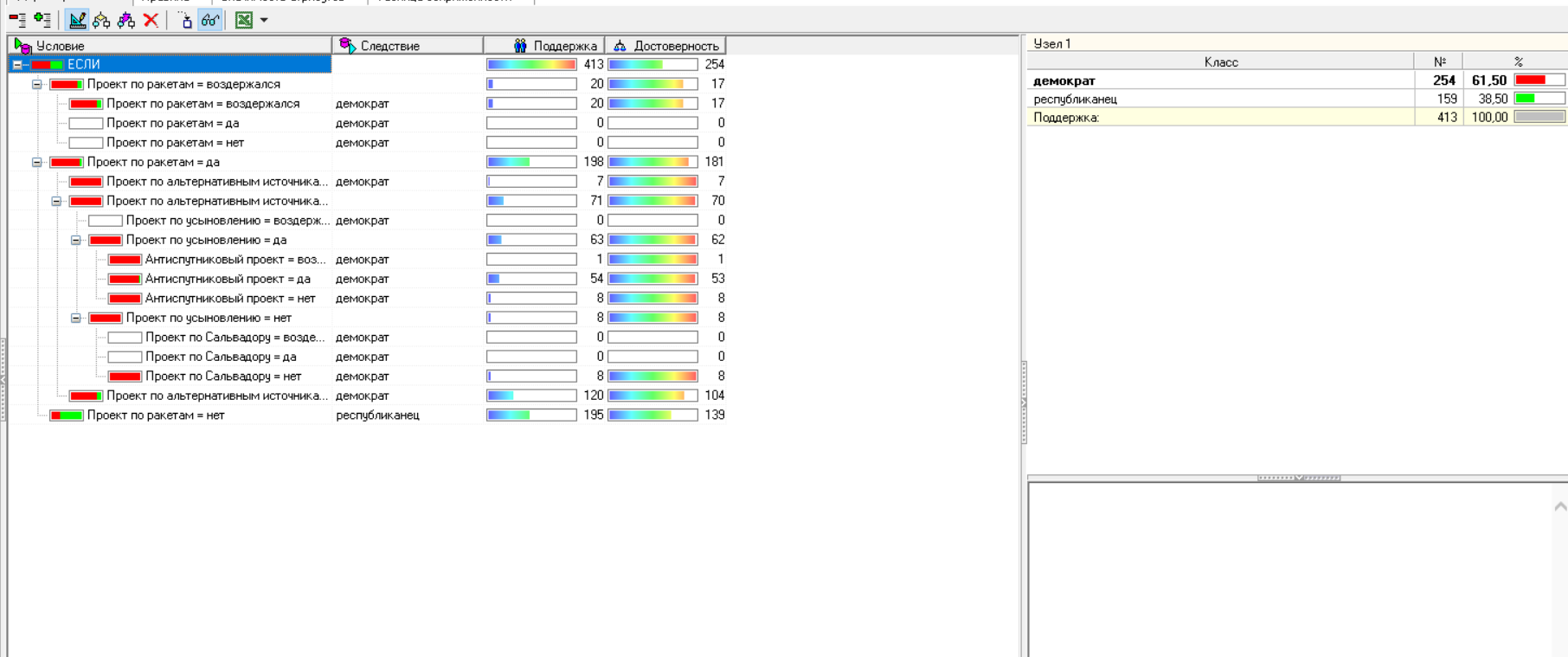
Для внесения изменений в него используют следующие кнопки:

включение/выключение интерактивного режима;

разбить текущий узел на подузлы;

построить дерево решений начиная с текущего узла.

Пусть аналитик думает, что основное правило, которое надо учитывать в построение дерева решений есть проект о ракетах. Тогда для данного построения выберем корневой каталог в дереве решений и нажмем кнопку и в появившемся окне выберем проект по ракетам. В результате получим новое дерево решений с новыми правилами и законами.



Выводы. Пример показал простоту и удобство применения деревьев решений для классификации на республиканцев и демократов.

Мастер предлагает широкие возможности по настройке процесса построения дерева решений: настройка назначения столбцов, способов нормализации, настройка источника данных для учителя (тестовое и обучающее множества), настройка количества примеров в узле и настройка достоверности правил.

Достоинства Дерева решений для анализа.

Алгоритм сам отсек несущественные факторы, выявил степень влияния тех или иных факторов на результат, описал при помощи формальных правил способ классификации, а также выдал информацию о достоверности и поддержке того или иного правила.

Также продемонстрированы широкие возможности визуализации построенного дерева.