

## АНАЛИЗ ДАННЫХ И ПОСТРОЕНИЕ ПРОГНОЗА В СИСТЕМЕ 1С:ПРЕДПРИТИЕ

В.В. Колмыков

**Аннотация.** Для эффективного прогнозирования необходимо понять механизм анализа данных. Данная статья познакомит с основными элементами механизма анализа данных и прогнозирования. Будут рассмотрены основные методы анализа данных. Показано на основе чего строится прогноз. Все методы и элементы анализа данных и прогнозирования будут рассмотрены в системе 1С:Предприятие с помощью обработки консоль анализа данных.

**Ключевые понятия:** анализ данных, прогнозирование, параметры, источник данных, тип анализа.

Механизм анализа данных и прогнозирования позволяет реализовать в прикладных решениях различные средства для выявления закономерностей, которые обычно скрываются за большими объемами информации.

Анализ данных – это способ нахождения неочевидных закономерностей в анализируемых данных и представление результата в легко доступной форме. Схема взаимодействия элементов механизма анализа данных и прогнозирования представлена на рис.1.

Анализ данных будем проводить с помощью консоли анализа данных рис.2. Даная обработка поставляется с системой 1С:Предприятие (ИТС).

Прогнозирование, разработка прогноза; в узком значении – специальное научное исследование конкретных перспектив развития какого-либо процесса

Механизм позволяет работать как с данными, полученными из информационной базы, так и с данными, полученными из другого источника, предварительно загруженными в таблицу значений или табличный документ.

В качестве источника данных может использоваться результат запроса, таблица значений, область ячеек табличного документа.

Применяя к исходным данным один из видов анализа, можно получить результат анализа. Результат анализа представляет собой некую модель проведения данных. Результат анализа может быть отображен в итоговом документе или сохранен для дальнейшего использования.

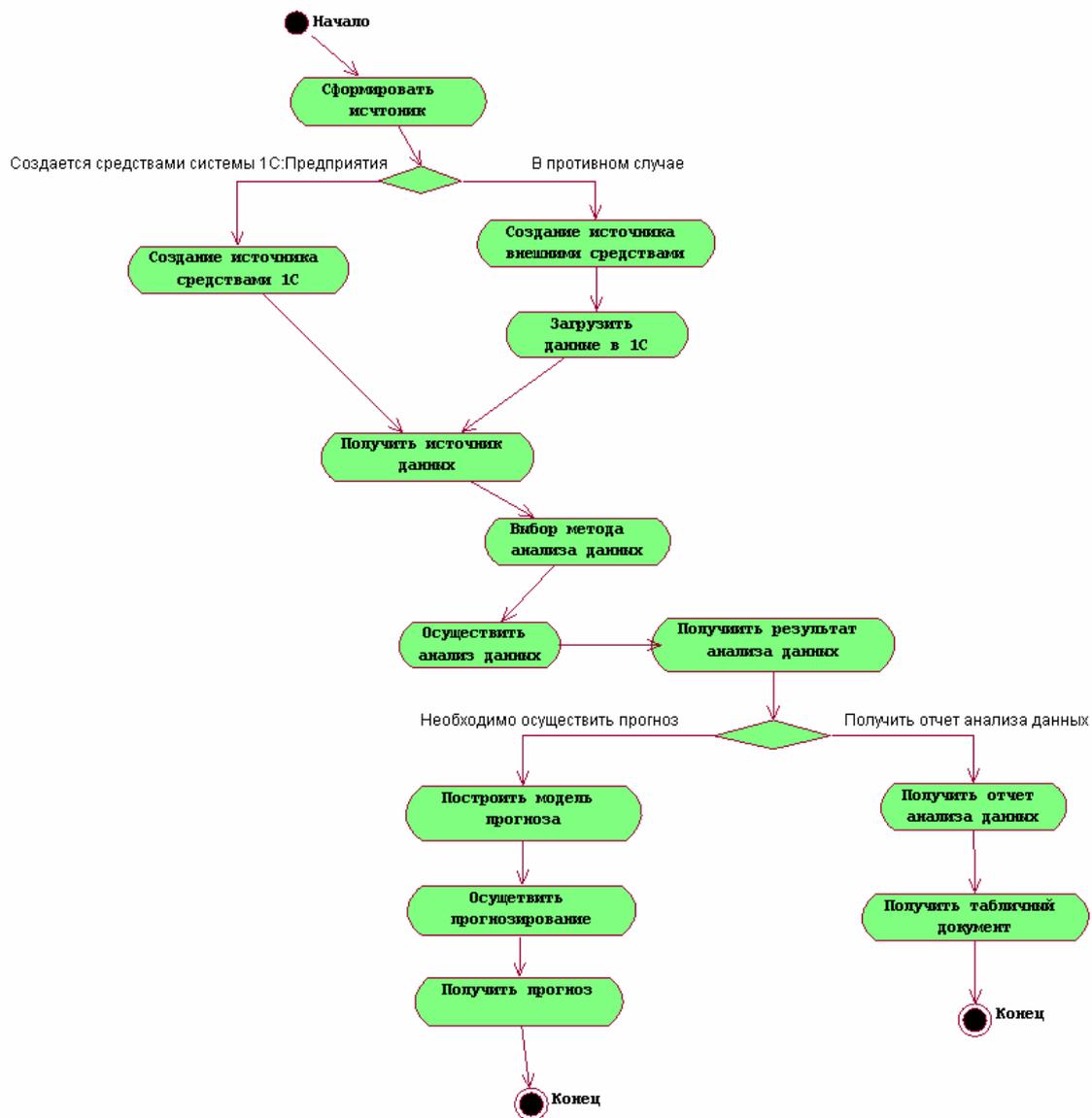


Рис.1. Схема взаимодействия элементов механизма анализа данных и прогнозирования

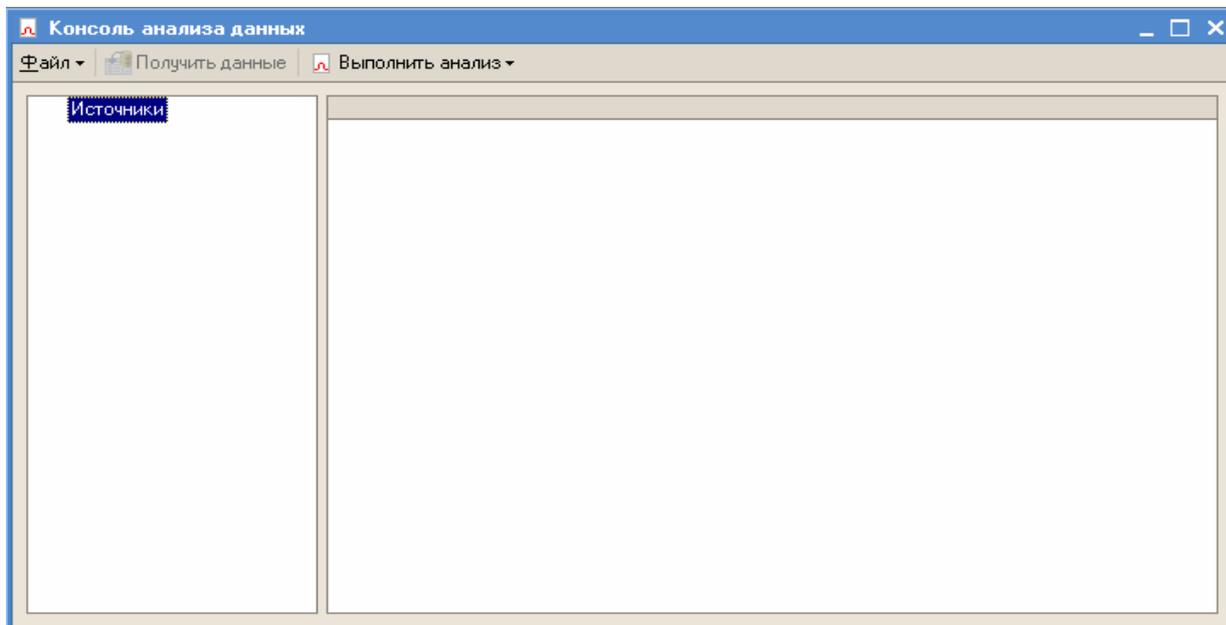


Рис.2. Консоль анализа данных

### **Тип анализа общая статистика.**

Тип анализа данных необходим для исследования источника информации. Позволяет получать информацию о данном источнике. Что позволит представить общую картину рассматриваемой проблемы.

Для колонок разных видов предусмотрено получение различной информации.

#### ***Дискретные данные:***

- Количество значений. Количество значений, встреченных в колонке источника данных (Null значением не считается);
- Количество уникальных значений;
- Мода. Значение, которое в источнике данных встречается наиболее часто. Если данных несколько значений, встречаемых с одинаковой частотой, в качестве моды берется первое найденное;
- Частота. Количество значений в выборку данных;
- Относительная частота. Определяется как отношение количества вхождения значения к общему количеству значений;
- Накопленная частота. Считается как сумма частоты значения и сумма частот предыдущих значений выборки данных;
- Накопленная относительная частота. Считается как сумма накопленной частоты значения и сумма относительных частот предыдущих значений;

#### ***Непрерывные данные:***

- Количество значений;
- Минимум значений;
- Максимум значения;
- Среднее;
- Размах. Разность между максимальным и минимальным значениями;

- Стандартное отклонение (среднеквадратичное отклонение);
- Медиана. Значение, лежащие в середине выборки.

Следует отметить, что если анализируется одновременно несколько полей различных видов, их анализ проводится вне зависимости друг от друга.

Результат проведения анализа общая статистика представлен на рис.3.

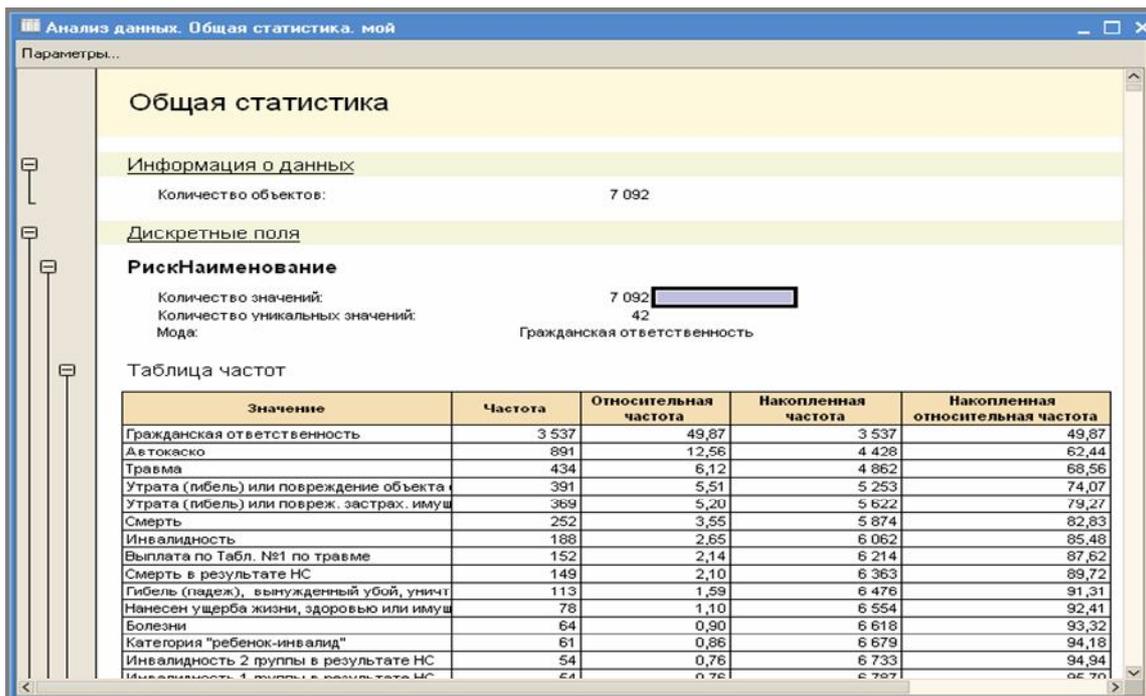


Рис.3. Анализ данных общая статистика

### **Тип анализа поиск ассоциаций**

Данный тип анализа осуществляет поиск часто встречаемых вместе комбинаций объектов или событий. С точки зрения данного типа анализа колонки источника можно разделить на:

- ✓ Неиспользуемая. Игнорируются анализом;
- ✓ Объект. Данные из этой колонки используются как объекты (или события) проводимого анализа. Исходя из значения данной колонки значения другой колонки (Элемент) относятся к одной ассоциируемой группе;
- ✓ Элемент. Данные из этой колонки используются для получения устойчивых значений, построения ассоциативных правил.

Рассмотрим параметры анализа рис.4:

- МинимальныйПроцентСлучаев. Определяется минимальный процент случаев, в которых должна встречаться комбинация элементов. Группы, у которых данное значение меньше указанного, не попадает в результат анализа;
- МинимальнаяДостоверность. Показывает минимальное значение процента случаев, когда правило соблюдается;
- МинимальнаяЗначимость. Группы, у которых данное значение меньше указанного, не попадают в результат анализа;

- ТипОтсеченияПравил. Один из вариантов системного перечисления ТипОтсеченияПравилАссоциации:Избыточные (отсекать избыточные правила), Покрытые (отсекать правила, покрытые другими правилами).

В результате выполнения анализа получаем рис.5:

- информация о данных (количество объектов, количество элементов, среднее количество элементов в объекте, количество найденных правил ассоциаций);
- найденные группы элементов. Указывается состав группы, количество случаев, процент случаев в которых эта группа встречается;
- найденные ассоциативные правила. Указывается исходный состав элементов, следствие (состав элементов), процент случаев, достоверность, значимость правила.

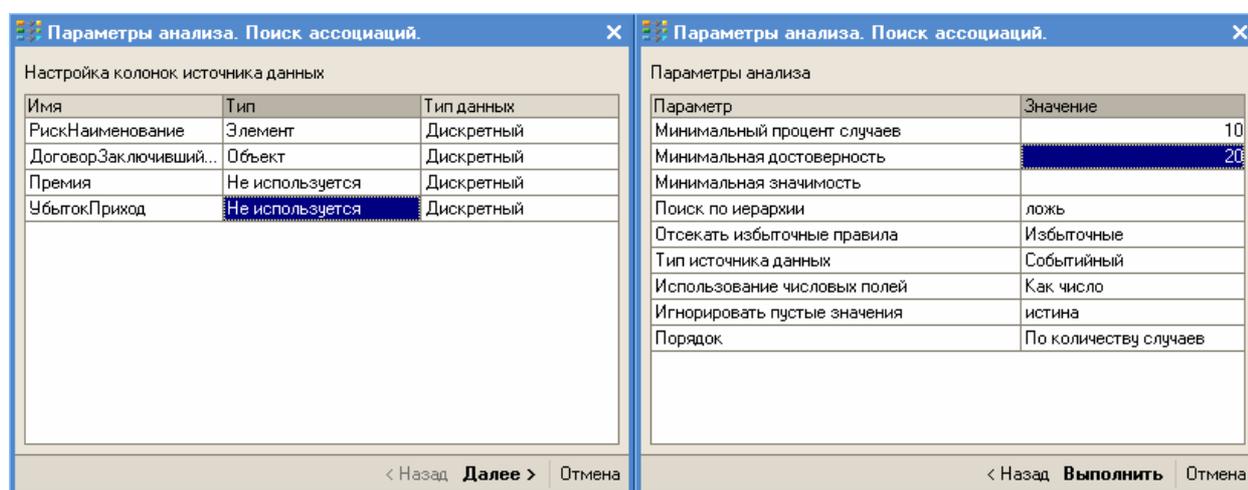


Рис. 4. Параметры анализа данных

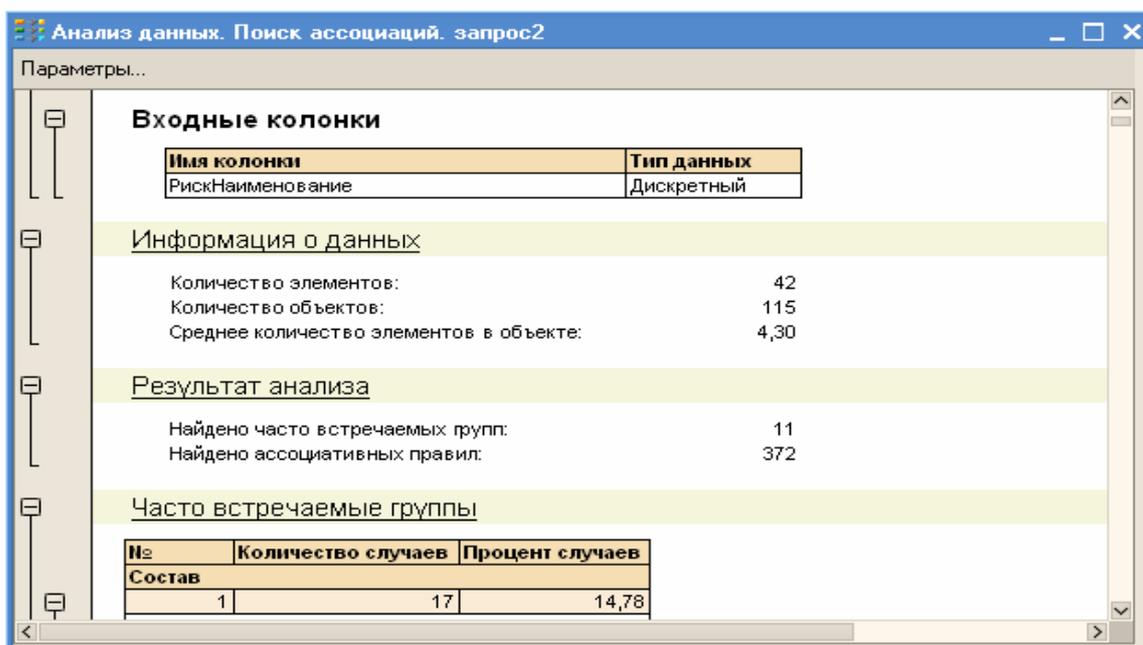


Рис. 5. Результат анализа поиска ассоциаций

## Поиск последовательностей

Анализ данного типа позволяет выявить цепочки возникающих событий. Он может использоваться тогда, когда одним из важных анализируемых показателей является последовательность товаров, которые закупаются друг за другом в течение какого – либо определенного промежутка времени.

С точки зрения данного типа анализа колонки источника представлены следующим образом:

- ✓ Неиспользуемая;
- ✓ Последовательность. Данные из этой колонки используется для анализа как объект события последовательности. По значению данной колонки анализа и ассоциирует данные с одной цепочкой событий;
- ✓ Элемент;
- ✓ Время. Именно по данной колонке определяется время наступления события. Наличие данной колонки обязательно при проведении данного типа анализа;

Кроме настройки типов колонок, на результат проводимого анализа влияют следующие параметры анализа:

- МинимальныйПроцентСлучаев.;
- МинимальныйИнтервал. Признак установки минимального интервала последовательности (должна быть установлена единица измерения интервала, кратность);
- МаксимальныйИнтервал. Признак установки максимального интервала последовательности (должна быть установлена единица измерения интервала, кратность);
- ИнтервалЭквивалентностиВремени. Признак установки интервала эквивалентности времени (должна быть установлена единица интервала эквивалентности времени, ее кратность);
- МинимальнаяДлина;
- ПоискПоИерархии. Признак осуществления поиска по иерархии (распространяется на колонки с типом Элемент)

Основным результатом анализа являются найденные шаблоны последовательностей. Этот шаблон содержит следующую информацию:

- состав шаблона последовательности;
- количество случаев, в котором наблюдалась данная последовательность;
- максимальные интервалы между событиями (в случае, если событий 2, то интервал один);
- минимальные интервалы между событиями (в случае, если событий 2, то интервал один);
- процент случаев, когда данная последовательность выполнялась;
- средние интервалы между событиями (в случае, если событий 2, то интервал один).

## Тип анализа дерево решений

С помощью данного типа анализа можно получить причинно-следственную иерархию условий, приводящую к определенному решению. Например, получить дерево условий, по которому (с определенной долей вероятности) можно понять причину расторжения договоров с клиентами компании, определения условий, влияющих на вариант заключаемого договора.

С точки зрения данного типа анализа колонки источника можно разделить на:

- ✓ Неиспользуемая;
- ✓ Входная;
- ✓ Прогнозируемая.

Используемые параметры анализа:

- Минимальное количество элементов в узле — минимальное количество элементов в узле;

- Максимальная глубина - максимальная глубина дерева;

- Тип упрощения - тип упрощения дерева решений.

И в результате проведенного анализа можно получить:

- дерево решений таблица №1;
- ошибки классификации таблица №2.

Ошибки классификации показывают, в каких случаях полученные правила расходятся с «действительностью» (исходной выборкой данных).

Упрощение дерева заключается в том, что по определенным правилам (формулам, которые будут рассмотрены ниже) узлы дерева превращаются в листья (отсекается «лишнее» ветвление).

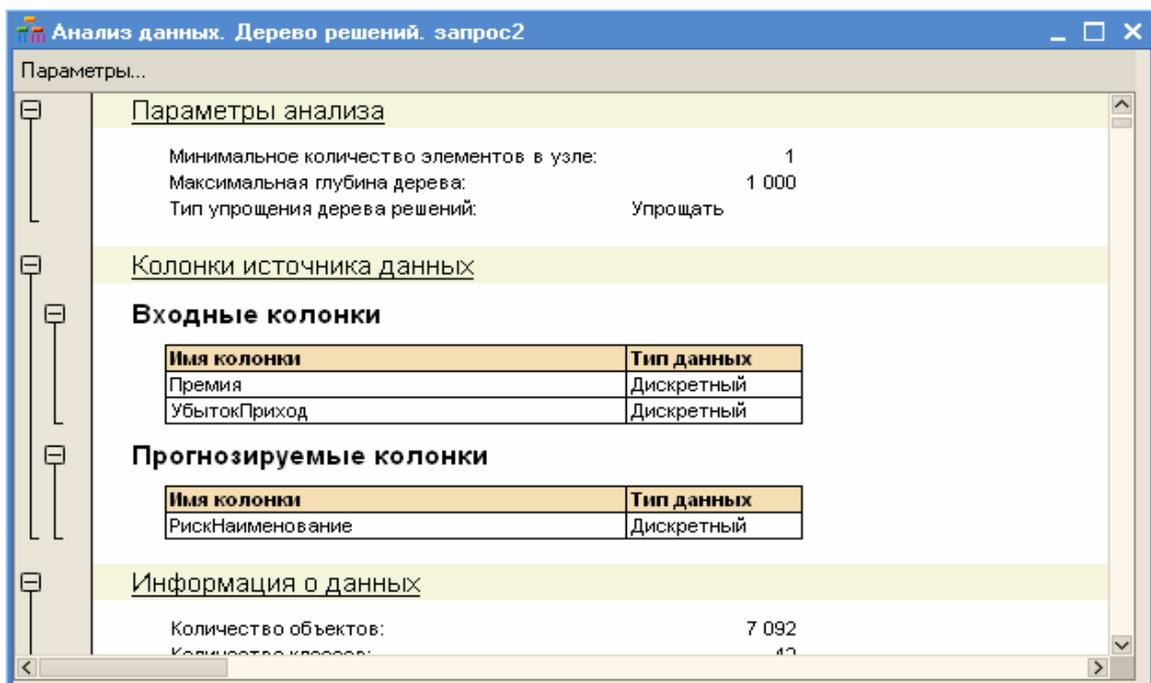


Рис. 6. Результат анализа поиска ассоциаций

Таблица №1 – Дерево решений

Условие	Решение	Вероятность	Количество
Корень	ДоговорЗаклучившийНаименование =	25,00	7 092
РискНаименование = Событ, повлекш за собой причинен вреда тр лицам	ДоговорЗаклучившийНаименование =	44,44	27
РискНаименование = АвтоДО	ДоговорЗаклучившийНаименование =	44,44	9
РискНаименование = Автокаска	ДоговорЗаклучившийНаименование =	43,21	891
РискНаименование = Болезни	ДоговорЗаклучившийНаименование =	28,13	64
РискНаименование = Возник-е обяз-ти Стр-ля возместить ущерб имущ-ву	ДоговорЗаклучившийНаименование =	100,00	1
РискНаименование = Все риски	ДоговорЗаклучившийНаименование =	100,00	1
РискНаименование = Выплата за каждый день лечения из-за травмы	ДоговорЗаклучившийНаименование =	100,00	1

Таблица №2 – Ошибка классификации

Факт \ Классифицировано	ДоговорЗаклучившийНаименование =			
	ДоговорЗаклучившийНаименование = Автокомплекс Зубово Поляна ООО			
	ДоговорЗаклучившийНаименование = Адмайкин Алексей Анатольевич			
	ДоговорЗаклучившийНаименование = Азаркина Татьяна Михайловна			
ДоговорЗаклучившийНаименование =	205			
ДоговорЗаклучившийНаименование = Автокомплекс				
ДоговорЗаклучившийНаименование = ФИО	4			
ДоговорЗаклучившийНаименование = ФИО	2			
ДоговорЗаклучившийНаименование = ФИО*				

\*Указывается фамилия имя и отчество заключившего договор

## **Тип анализа кластеризация**

Кластеризация — выделение из множества объектов одной природы некоторого количества относительно однородных групп — сегментов или кластеров. Объекты распределяются по группам так, чтобы внутригрупповые отличия были минимальными, а межгрупповые — максимальными. Методы кластеризации позволяют перейти от пообъектного к групповому представлению совокупности произвольных объектов, что существенно упрощает оперирование ими.

В основе данного анализа лежит вычисление расстояния между группами, которые и являются кластерами. Определение расстояния между группами производится различными способами (по разным метрикам). Поддерживаются следующие метрики:

- Евклидова метрика;
- Евклидова метрика в квадрате;
- Метрика города;
- Метрика доминирования.

После определения расстояний между объектами может использоваться один из нескольких алгоритмов распределения объектов по кластерам. Поддерживаются следующие методы кластеризации:

- Ближняя связь;
- Дальняя связь;
- Центр тяжести;
- «к средних».

Колонки источника определяются как входные либо не используемые. Анализ производится в соответствии с установленными параметрами анализа.

Результатами анализа рис. 7 с помощью кластеризации являются:

- центры кластеров, представляющие собой совокупность усредненных значений входных колонок в каждом кластере;
- таблица межкластерных расстояний (между центрами кластеров), определяющих степень различия между ними;
- значения прогнозных колонок для каждого кластера;
- рейтинг факторов и дерево условий, определивших распределение объектов на кластеры.

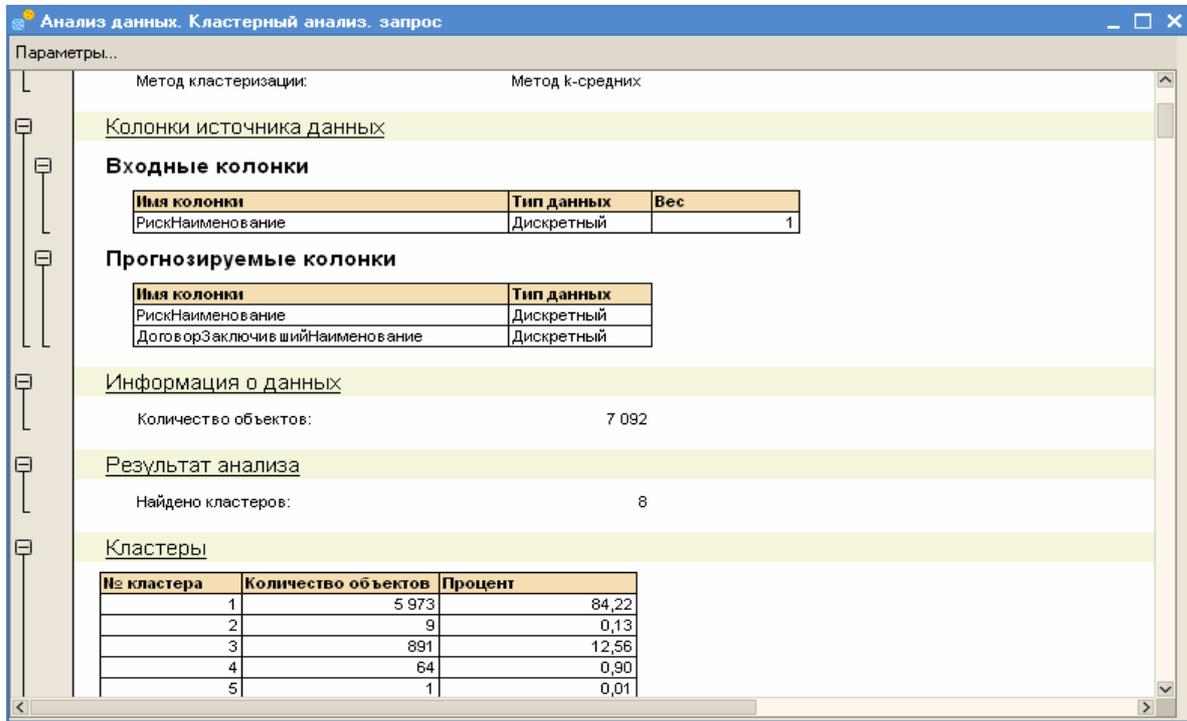


Рис. 7. Результат анализа кластеризация

## Проведение прогнозов

Для проведения прогнозирования необходимо осуществить следующие этапы:

- обучение модели исходя из выбранных данных;
- использование обученной модели для осуществления прогнозирования.

Исходя, из жизненного опыта можно сказать, что перед тем как принять какое-то решение относительно прогноза человек (эксперт) длительное время изучает закономерности, учиться правильно решать.

В системе один ИС:Предприятия в качестве такого эксперта выступает модель прогноза. Данный объект получается на основе обучения модели, то есть модель, полученная на основе анализа данных листинг 1. Прогноз рассмотрен на основе кластерного анализа. Результат проведения прогноза представлен в таблице №3

### Листинг 1.

```

Анализ=Новый АнализДанных;
Анализ.ТипАнализа = Тип("АнализДанныхКластеризация");
Анализ.ИсточникДанных = ТекущийИсточникДанных;
Анализ.НастройкаКолонок.РискНаименование.ТипКолонки=ТипКолонкиАнализаДанныхКластеризаци
я.ВходнаяИПрогнозируемая;
Анализ.НастройкаКолонок.ДоговорЗаклучившийНаименование.ТипКолонки=ТипКолонкиАнализаДан
ныхКластеризация.Прогнозируемая;
Анализ.НастройкаКолонок.Премия.ТипКолонки=ТипКолонкиАнализаДанныхКластеризация.Ключ;
Анализ.Параметры.КоличествоКластеров.значение=8;
Анализ.Параметры.мераРасстояния.значение=ТипМерыРасстоянияАнализаДанных.ЕвклидоваМетрика;
Анализ.Параметры.методКластеризации.значение=методкластеризации.КСредних;
МодельПрогноза=РезультатАнализа.СоздатьМодельПрогноза();
МодельПрогноза.ИсточникДанных=ТекущийИсточникДанных;

```

ТаблицаЗначенийПоследовательностей=МодельПрогноза.Выполнить();

Если ТаблицаЗначенийПоследовательностей.Количество() $>0$  тогда

РезультатПрогноза.Очистить();

ЭлементыФормы.РезультатПрогноза.Значение=ТаблицаЗначенийПоследовательностей;

ЭлементыФормы.РезультатПрогноза.СоздатьКолонки();

КонецЕсли;

Если РезультатПрогноза.Количество() $=0$  тогда

Предупреждение("Предложений нет");

Иначе

ЭлементыФормы.РезультатПрогноза.СоздатьКолонки();

Создание модели прогноза доступно для всех типов анализа за исключением общей статистики.

Таблица 3 – Результат прогноза

Риск	Премия	Кластер	РискНаименование	РискВероятность	ЗаключившийВероятность
Событ, повлекш за собой причинен вреда тр лицам	800,00	1	Гражданская ответственность	57,1	22,4
Событ, повлекш за собой причинен вреда тр лицам	400,00	1	Гражданская ответственность	57,1	22,4
АвтоДО	828,00	2	АвтоДО	100	44,4
Автокаско	10 240,00	3	Автокаско	100	43,1
Автокаско	38 902,00	3	Автокаско	100	43,1

### Список использованной литературы

1. Радченко, М.Г. 1С:Предприятие 8.1. Практическое пособие разработчика/ М.Г. Радченко. П., 2007 с.357
2. Габец ,А.П., Гончаров, Д.И., Козырев, Д.В. Профессиональная разработка в системе 1С:Предприятие 8/ А.П. Габец, Д.И. Гончаров, Д.В. Козырев., 2007 с.398

### Сведения об авторе

Колмыков Вячеслав Викторович, аспирант кафедры «Автоматизированные системы обработки информации и управления»