

**Лауреат Конкурса-2015  
В номинации «Развитие финансовой системы»**

**Сафиуллин Руслан Фаритович**

Академия ВЭГУ

Научный руководитель: Шакирова Ф.М.

**«Использование интеллектуальных информационных систем  
для анализа данных банковского сектора»**

**Введение**

Согласно обзору аналитического агентства TAdviser на российском рынке ППР (поддержки принятия решений) или DSS (decision support system), в финансовой сфере наиболее популярными являются программы Прогноз (Прогноз, Рязань), QlikView (QlikTech, Sweden) и Deductor (BaseGroup, Рязань). С сайта компании BaseGroup была скачана и в компьютерных классах установлена свободно распространяемая учебная версия аналитической платформы Deductor Academic, предназначенной для задач интеллектуальной обработки данных. В компьютерных классах Академии ВЭГУ также установлено лицензионное программное обеспечение, купленное у компании Microsoft по программе MSDN Academic Alliance, в частности, Microsoft Excel 2003 и Microsoft Excel 2007.

Наша цель – проверка возможностей программ для решения аналитических задач, касающихся банковской деятельности.

План намеченных работ:

1. Поиск источников для реализации цели.
2. Использование указанных программ для решения задач Data Mining.

## 1. Источники данных

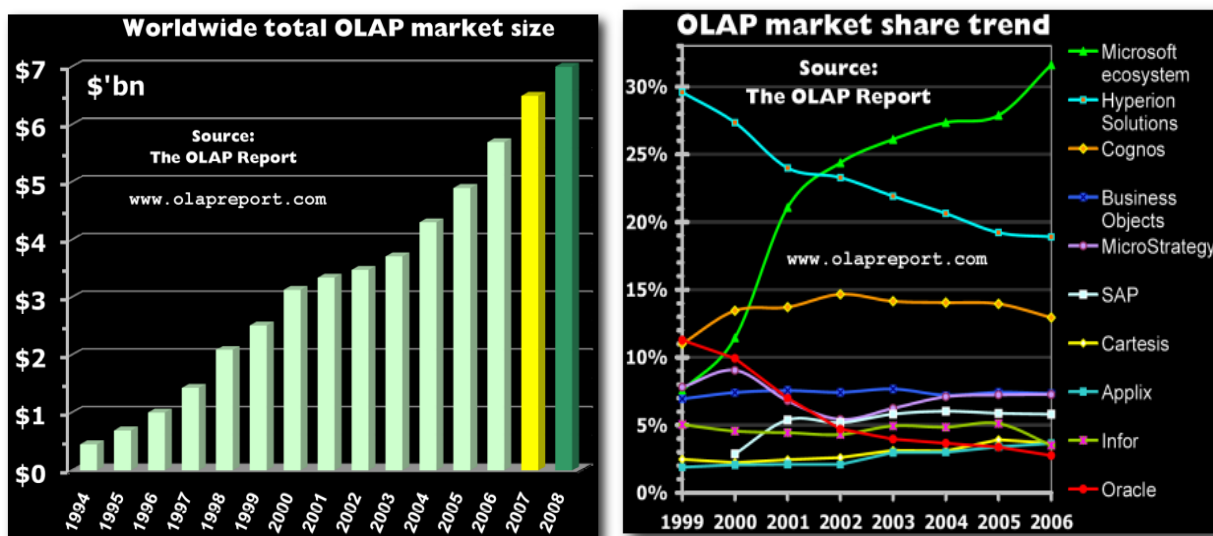
В качестве исходных данных использованы файлы:

- **bnkseek.dbf** со справочником банковских идентификационных кодов (БИК) кредитных организаций (КО) России. Файл распространяется Банком России в свободном доступе в сети Интернет (например, по адресу <http://www.pvision.ru/bik>) и будет использован нами для создания на его основе сводной таблицы, отражающей количество КО в Уральско-Поволжском регионе (для сокращения времени реализации задачи возьмем пять областей).
- **статистические данные** по выдаче кредитов виртуальным банком (файл поставляется вместе с учебной версией программы Deductor)

## 2. Создание сводной таблицы средствами MS Excel

### 2.1 Выборка банков из справочника БИК

По данным журнала “The OLAP Report” ([www.olapreport.com](http://www.olapreport.com), в настоящее время ссылка перенаправляется на сайт разработчика BI-систем) рынок OLAP-систем растет и наибольшую долю рынка имеет корпорация Microsoft. Данный вывод сделан на основе исследования компании Green Hill Analysis по использованию программ, реализующих технологию OLAP. Ниже представлены диаграммы компании



По данным исследовательской компании у Microsoft больше всего

драйверов для подключения к внешним данным

Microsoft	Oracle	IBM	Hyperion
Microsoft SQL Server Oracle ODBC Files Access 2000, Excel 2000 Microsoft Visual FoxPro dBase, Paradox Microsoft Exchange Server Microsoft Active Directory	IBM DB2 Informix Microsoft SQL Server Sybase Oracle ODBC Files	DB2 Informix Microsoft SQL Server Oracle Sybase Files	IBM DB2 Informix Microsoft SQL/Server Oracle Sybase ODBC

По признанию пользователей, достоинством MS Excel является концепция многомерных настольных таблиц и инструмент “сводные таблицы” в составе MS Excel. Это единственный факт широкого использования возможностей многомерного анализа, т.к. в мире много пользователей MS Excel.

Справочник банковских идентификационных кодов БИК (bnkseek.dbf) формируется Банком России. Он используется для автоматизированного заполнения банковскими атрибутами финансовых документов. Это единственное применение содержащейся в файле информации. Справочник еженедельно обновляется, что позволяет проследить изменения в состоянии банковской сферы России (в файле отсутствуют записи о банках, у которых отозвана лицензия).

Структура наших исследований:

1. распределение количества КО в РБ по годам выдачи лицензий;
2. анализ тенденций в численности местных банков и филиалов инорегиональных банков на территории РБ и соседних регионов;
3. интерпретация полученных результатов.

Вследствие того, что на сайте Банка России представлены данные до 2007 года и с ноября 2014 года, воспользуемся имеющимися в нашем распоряжении данными за прошлые годы, в частности по состоянию на 02.03.2009. Используем поля таблицы: признак банка (20 – головное учреждение, 30 – филиал), регион России, город, банк, дата получения лицензии. Фрагмент справочника:

	A	B	C	D	E
1	<b>признак</b>	<b>регион</b>	<b>город</b>	<b>банк</b>	<b>дата</b>
2	40	00	СЕВАСТОПОЛЬ	ПУ БАНКА РОССИИ N 10462	20.01.1994
3	40	00		ПУ БАНКА РОССИИ N 03148	20.01.1994
4	40	00		ПУ БАНКА РОССИИ N 83604	17.02.1994
5	40	00		ПУ БАНКА РОССИИ N 83524	20.01.1994
6	40	00		ПУ БАНКА РОССИИ N 25631	20.01.1994
7	40	00	БИШКЕК	ПУ БАНКА РОССИИ N 10513	20.01.1994
8	40	00		ПУ БАНКА РОССИИ N 18001	20.01.1994
9	40	00		ПУ БАНКА РОССИИ N 64106	20.01.1994
10	10	01	СОЛОНЕШНОЕ	РКЦ СОЛОНЕШНОЕ	20.01.1994
11	10	01	БЫСТРЫЙ ИСТОК	РКЦ БЫСТРЫЙ ИСТОК	20.01.1994
12	10	01	СОЛТОН	РКЦ СОЛТОН	20.01.1994
13	10	01	ЗМЕИНОГОРСК	РКЦ ЗМЕИНОГОРСК	20.01.1994

### Свод кредитных организаций по городам РБ

РЕГИОН	80												
Количество по ГОРОД													
ДАТА	БЕЛОРЕЦК	БИРСК	МЕЖГОРЬЕ-1	МЕСЯГУТОВО	НЕФТЕКАМСК	ОКтябрьский	СИБАЙ	СТЕРЛИТАМАК	ТУЙМАЗЫ	УФА	УЧАЛЫ	Общий итог	
1994		1	1	1	1	1	1		2	1	13	1	23
1999											1		1
2003											2		2
2004											2		2
2005								1			4		5
2006							1				11		12
2007											6		6
2008											8		8
Общий итог		1	1	1	1	1	1	1	3	1	47	1	59

На указанную дату в РБ действовало 59 КО, из них в Уфе было сосредоточено 47 КО. Наибольшее количество КО, имеющих на указанную дату лицензии, было зарегистрировано в 1994 году (23). Из тех КО, что были зарегистрированы на территории РБ в 1999, лицензии на банковскую деятельность оставались у 1 КО (филиал ОАО «ГАЗПРОМБАНК»).

1. Наибольшее количество банков, сохранивших лицензию, получили их в 2006 году.

### Свод КО по соседним регионам

Количество по полю БАНК		
РЕГИОН	Итог	
80	59	<b>Башкортостан</b>
57	76	<b>Пермская обл</b>
92	87	<b>Татарстан</b>
75	89	<b>Челябинская обл</b>
65	96	<b>Свердловская обл</b>
Общий итог	407	

### Количество лицензий по годам

РЕГИОН	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Общий итог
80	23					1				2	2	5	12	6	8		59
57	24	1		2	3	6	3		2	4	2	3	6	12	8		76
92	44	3	1	2	3	4	2	1	4	1	4	2	4	6	5	1	87
75	29	2	4		1	1	1	5		4	7	3	5	11	16		89
65	42	3	1	2	1	4	3	2	1	1	1	8	8	12	7		96
Общий итог	162	9	6	6	8	16	9	8	7	12	16	21	35	47	44	1	407

Меньше всего осталось банков, получивших лицензии в 2001-2003. Стабильная ситуация по всем годам прослеживалась лишь в Татарстане.

Посмотрим свод по названиям КО. Филиалы ЗАО "РАЙФФАЙЗЕНБАНК" и ОАО "АЛЬФАБАНК" называются "БАШКОРТОСТАН". В 2005 году БАШКРЕДИТБАНК стал филиалом ОАО "УРАЛСИБ", что добавило два "новых" банка

БАШКОРТОСТАН	ФИЛИАЛ "БАШКОРТОСТАН" ЗАО "РАЙФФАЙЗЕНБАНК"	2008	УФА
	ФИЛИАЛ "БАШКОРТОСТАН" ЗАО КБ "СИТИБАНК"	2008	УФА
	ФИЛИАЛ "БАШКОРТОСТАН" ОАО "АЛЬФА-БАНК"	2006	УФА
УРАЛСИБ	ФИЛИАЛ ОАО "УРАЛСИБ" В Г. СТЕРЛИТАМАК	2005	СТЕРЛИТАМАК
	ФИЛИАЛ ОАО "УРАЛСИБ" В Г.УФА	2005	УФА

Из-за разночтения в названиях КО необходимо выявлять и учитывать эту информацию.

По данным прошлых лет составим свод на начало года

23.03.2007 Было 4483 КО																		
Количество по полю NAM	DAT																	
RGN	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	Общий итог			
80	28					1				2	2	5	13	1	52		Башкортостан	
57	30	1	1	2	3	6	3		2	4	2	3	7	4	68		Пермская обл.	
75	39	2	4		1	2	3	5	1	6	8	3	5	4	83		Челябинская обл.	
92	47	3	2	2	3	5	3	1	4	1	4	2	4	4	85		Татарстан	
65	50	3	1	2	1	4	3	3	1	1	1	9	8	4	91		Свердловская обл.	
Общий итог	194	9	8	6	8	18	12	9	8	14	17	22	37	17	379			
05.10.2007 Было 4499 КО																		
Количество по полю NAM	DAT																	
RGN	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	Общий итог			
80	25					1				2	2	5	12	3	50			
57	28	1	1	2	3	6	3		2	4	2	3	7	11	73			
75	33	2	4		1	2	3	5		6	8	3	5	12	84			
92	46	3	1	2	3	4	3	1	4	1	4	2	4	6	84			
65	47	3	1	2	1	4	3	3	1	1	1	9	8	10	94			
Общий итог	179	9	7	6	8	17	12	9	7	14	17	22	36	42	385			
23.01.2009 и 02.03.2009 Было 4404																		
Количество по полю NAM	DAT																	
RGN	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Общий итог	
80	23					1				2	2	5	12	6	8		59	
57	24	1		2	3	6	3		2	4	2	3	6	12	8		76	
92	44	3	1	2	3	4	2	1	4	1	4	2	4	6	5	1	87	
75	29	2	4		1	1	1	5		4	7	3	5	11	16		89	
65	42	3	1	2	1	4	3	2	1	1	1	8	8	12	7		96	
Общий итог	162	9	6	6	8	16	9	8	7	12	16	21	35	47	44	1	407	

Наблюдается тенденция отзыва лицензий у КО, получивших их в 1994 году (194,179,162), главным образом, из-за закрытия РКЦ. На 10.2007 в России было 723 РКЦ, на 03.2009 – 543 РКЦ.

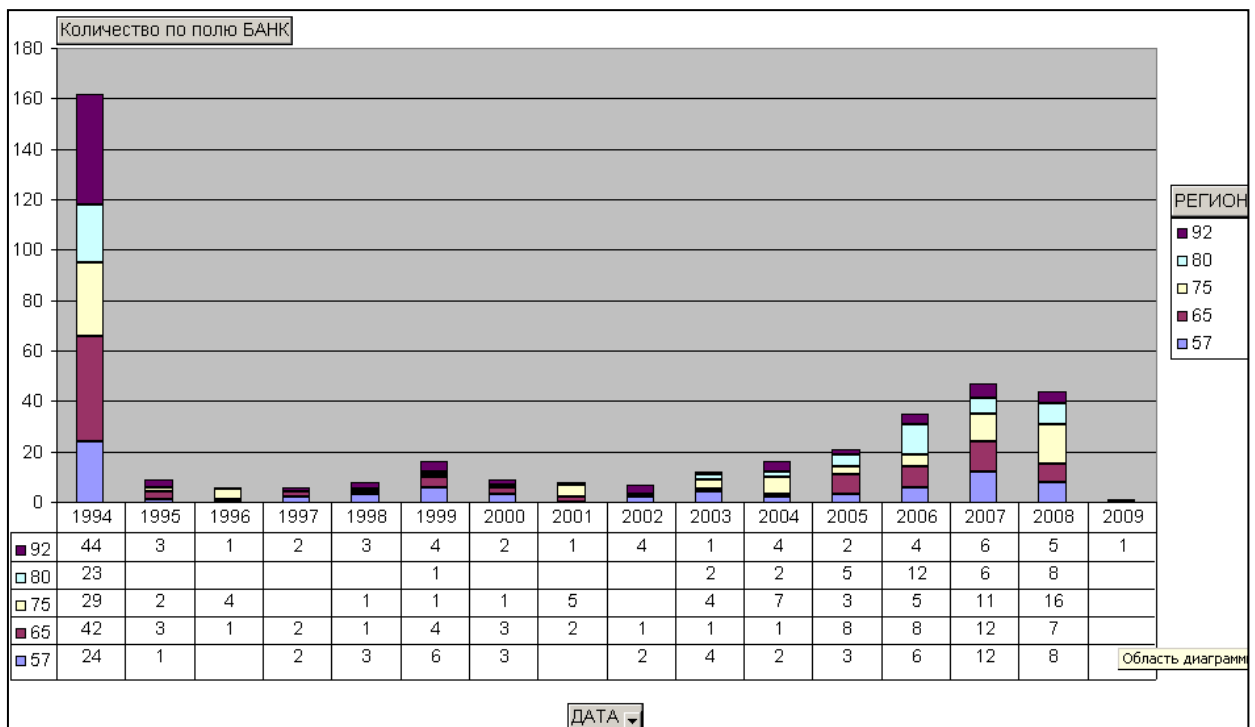
С экономическим кризисом банковская активность снизилась (в начале 2007 года лицензии получили 17 КО, в начале 2009 года – 1 КО).

Соотношение местных и инорегиональных КО

РЕГИОН	ПРИЗНАК	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Общий итог	
80	20	11																11	
	30							1			2	2	5	12	6	8		36	3,3
80 Итого		11						1			2	2	5	12	6	8		47	Башкортостан
57	20	8																8	
	30	3	1		2	3	6	3		2	4	2	3	6	11	8		54	6,8
57 Итого		11	1		2	3	6	3		2	4	2	3	6	11	8		62	Перская обл
92	20	23	1									1						25	
	30	7	1	1	2	2	4	2	1	4	1	3	2	4	6	5	1	46	1,8
92 Итого		30	2	1	2	2	4	2	1	4	1	4	2	4	6	5	1	71	Татарстан
65	20	21	2															23	
	30	4	1	1	1		4	3	2	1	1	1	8	7	12	6		52	2,3
65 Итого		25	3	1	1		4	3	2	1	1	1	8	7	12	6		75	Челябинская обл
75	20	9	1	1														11	
	30	8	1	2		1	1	1	5		4	7	3	5	11	16		65	5,9
75 Итого		17	2	3		1	1	1	5		4	7	3	5	11	16		76	Свердловская обл
Общий итог		94	8	5	5	6	16	9	8	7	12	16	21	34	46	43	1	331	

Наибольшая доля инорегиональных КО (признак банка - 30) в Пермской (в 6,8 раз) и Свердловской областях (в 5,9 раз), наименьшее количество – в республике Татарстан. В республике Башкортостан инорегиональных КО (36) в 3,3 раза больше, чем “местных” КО (11). По справочнику БИК на 16.10.2015 в РБ осталось 7 местных КО. Таким образом, в региональных округах происходит сокращение местных банков при общей тенденции сокращения банков России. На сайте [www.cbr.ru](http://www.cbr.ru) представлена статистика по численности банков до 01.09.2015 года, когда их стало 774.

“Гистограмма с накоплением”, позволяющая обобщить результаты



Интерпретация результатов наличия на рынке КО, получивших лицензии в разные годы и влияние внешних воздействий на их ликвидность:

- 1994г. - появление значительного числа КО из-за высокой инфляции, благодаря которой банки получали прибыль без реального развития квалифицированного управления. Налоги на доходы банков и страховщиков с 1994г. были отменены, они стали платить налог на прибыль, как и другие юрлица.
- 1998г. - экономический спад, связанный с резким падением цен на нефть.
- 2001 – 2002 гг. – менее значительный спад, он также вызван падением цен на сырьевые ресурсы.
- 2003 – 2004 гг. - создание системы гарантирования или страхования вкладов: с 01.01.2003 ЦБ РФ выдает новые лицензии на работу со средствами физических лиц только банкам, вступившим в ССВ.
- с 01.01.2005 — завершение переходного периода, отзыв лицензий на работу со средствами физических лиц у банков, не вошедших в ССВ.
- на 01.01.2005 в РБ из-за финансовой несостоятельности и нарушений банковского законодательства лицензии лишились 18 КО.

Выводы:

Мы обратились к известному справочнику для решения с его помощью нетривиальной задачи - добычи информации о состоянии банковского сектора в республике и в соседних регионах. В выпусках информационно-аналитического бюллетеня “Вестник Национального Банка Республики Башкортостан” (<http://www.finance.bashinform.ru>) отсутствуют подобного рода обзоры.

Учитывая полноту данных и периодичность обновления Банком России справочника БИК, он может быть использован для актуализации информации, а также для ретроспективного обзора развития банковской деятельности в любом регионе России.

## ***2.2 Использование программы Deductor для анализа данных***

Построим в Microsoft Excel сводную таблицу по справочнику БИК на 05.04.2015:

	A	B	C	D	E	F	G	H
1	PZN	(Все)						
2	RGN	80						
3								
4	Количество	NNP						
5	DATE_IN	БЕЛОРЕЦК	МЕЖГОРЬЕ-1	НЕФТЕКАМСК	СТЕРЛИТАМ	ТАМ	ТУЙМАЗЫ	УФА
6	1994	1	1	1	1	1	9	14
7	1999						1	1
8	2004						1	1
9	2005						3	3
10	2006						4	4
11	2008						1	1
12	2009						2	2
13	2010						1	1
14	2011						2	2
15	2013						14	14
16	2014						2	2
17	2015						1	1
18	Общий итог	1	1	1	1	1	41	46

Результаты обработки данных необходимо проверять. Загрузим исходные данные в Deductor Academic, сохранив предварительно справочник в текстовом файле (ограничение учебной версии программы). Выполним преобразование даты в представление “Год, Месяц” и построим куб OLAP

PZN		NNP						
- + D...		БЕЛОРЕЦК	МЕЖГОРЬЕ	НЕФТЕКАМ	СТЕРЛИТАМ	ТУЙМАЗЫ	УФА	Итого:
+ 1994	<input checked="" type="checkbox"/> 80				1	1	9	14
+ 1999	<input type="checkbox"/> 81						1	1
+ 2004	<input type="checkbox"/> 82						1	1
+ 2005	<input type="checkbox"/> 83						3	3
+ 2006	<input type="checkbox"/> 84						4	4
+ 2008	<input type="checkbox"/> 85						1	1
+ 2009	<input type="checkbox"/> 86						2	2
+ 2010	<input type="checkbox"/> 87						1	1
+ 2011	<input type="checkbox"/> 88						2	2
+ 2013	<input type="checkbox"/> 89						14	14
+ 2014	<input type="checkbox"/> 90						2	2
+ 2015	<input type="checkbox"/> 91						1	1
<b>Итого:</b>		1	1	1	1	1	41	46

Итоги совпадают.

Создадим текстовый файл, объединив данные справочников БИК на 06.04.2015 и на 16.10.2015. Для различия одинаковых записей введем дополнительный столбец с числом 1 в записях первого набора и числом 2 в записях второго набора. Загрузим файл в Deductor Academic и выполним обработку “Выявление дубликатов и противоречий”.

Отсортировав записи по датам, увидим противоречия в кратких названиях КО в справочниках БИК, а также банки, которые на 16.10.2015



лишились лицензии (их названия встречаются в 1-м файле и отсутствуют во 2-м файле, для проверки также можно использовать файл, содержащий список КО с отозванной лицензией)

Дубликаты		Противоречия		Ид. тол.	Входные поля		Выходные поля				Информационные поля	
Признак	группа	Признак	группа	file	NAMEN	DATE_IN (Год + Месяц)	NNP	PZN	RGN	DATE_IN		
<input type="checkbox"/>		<input type="checkbox"/>		1	УРАЛСИБ	01.09.2005	УФА	30	80	12.09.2005		
<input type="checkbox"/>		<input type="checkbox"/>		2	БАНК УРАЛСИБ	01.09.2005	УФА	30	80	12.09.2005		
<input type="checkbox"/>		<input type="checkbox"/>		1	БАШКИРСКИЙ ФИЛИАЛ ЗАО ЮНИКРЕДИ	01.12.2008	УФА	30	80	25.12.2008		
<input type="checkbox"/>		<input type="checkbox"/>		2	БАШКИРСКИЙ ФИЛИАЛ АО ЮНИКРЕДИТ	01.12.2008	УФА	30	80	25.12.2008		
<input type="checkbox"/>		<input type="checkbox"/>		1	МЕЖДУНАРОДНЫЙ БАНК РАЗВИТИЯ	01.12.2010	УФА	30	80	03.12.2010		
<input type="checkbox"/>		<input type="checkbox"/>		1	ИНКАХРАН	01.08.2013	УФА	74	80	23.08.2013		
<input type="checkbox"/>		<input type="checkbox"/>		1	ФИЛИАЛ N 20 АКБ МОСОБЛБАНК ОАО	01.08.2013	УФА	30	80	28.08.2013		
<input type="checkbox"/>		<input type="checkbox"/>		2	ФИЛИАЛ N 20 ПАО МОСОБЛБАНК	01.08.2013	УФА	30	80	28.08.2013		

Установив фильтр по признаку “Противоречия”, увидим запись о новом банке “ФК Открытие” на 16.10.2015

Дубликаты		Противоречия		Ид. тол.	Входные поля		Выходные поля				Информационные поля	
Признак	группа	Признак	группа	file	NAMEN	DATE_IN (Год + Месяц)	NNP	PZN	RGN	DATE_IN		
<input checked="" type="checkbox"/>	878	<input checked="" type="checkbox"/>	81	1	РОСТ БАНК	01.07.2013	УФА	30	80	17.07.2013		
<input checked="" type="checkbox"/>	878	<input checked="" type="checkbox"/>	81	2	РОСТ БАНК	01.07.2013	УФА	30	80	17.07.2013		
<input checked="" type="checkbox"/>	190	<input checked="" type="checkbox"/>	102	1	СОЦИНВЕСТБАНК	01.01.1994	УФА	20	80	20.01.1994		
<input checked="" type="checkbox"/>	190	<input checked="" type="checkbox"/>	102	2	СОЦИНВЕСТБАНК	01.01.1994	УФА	20	80	20.01.1994		
<input type="checkbox"/>		<input checked="" type="checkbox"/>	110	2	ФК ОТКРЫТИЕ	01.07.2015	УФА	30	80	09.07.2015		

Можно также посмотреть на результаты установки фильтра по признаку “Дубликаты” для выявления банков, сохранивших лицензии на 16.10.2015.

Число “местных” банков в РБ - 7

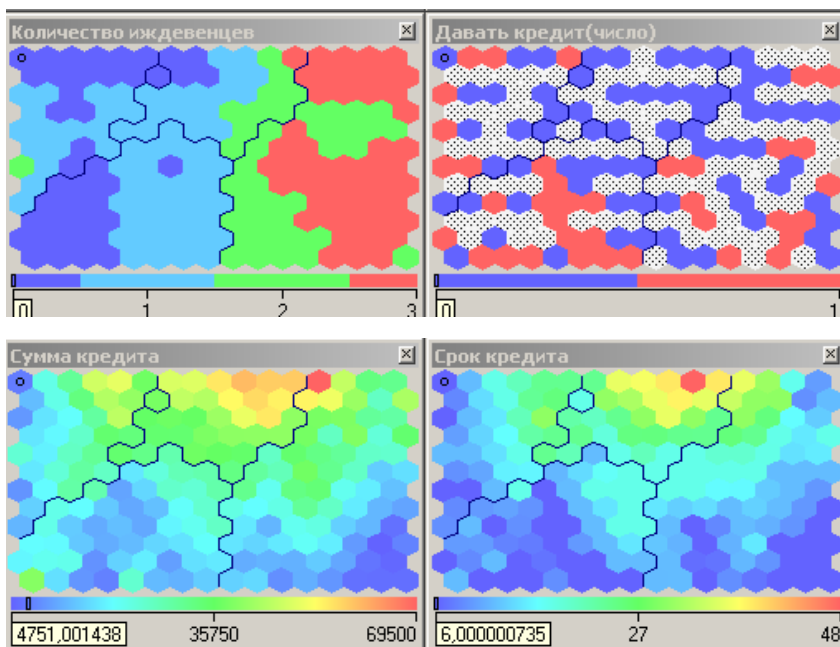
PZN	RGN	NNP	NAMEN	DATE_IN	DATE_IN (Год + Месяц)	file
20	80	УФА	БАНК ПТБ (ООО)	20.09.1994	01.09.1994	2
20	80	УФА	БАШКОМСНАББАНК (ПАО)	20.01.1994	01.01.1994	2
20	80	УФА	БАШПРОМБАНК	20.01.1994	01.01.1994	2
20	80	УФА	ИНВЕСТКАПИТАЛБАНК	20.01.1994	01.01.1994	2
20	80	УФА	РБР	20.09.1994	01.09.1994	2
20	80	УФА	СОЦИНВЕСТБАНК	20.01.1994	01.01.1994	2
20	80	УФА	УРАЛКАПИТАЛБАНК	17.02.1994	01.02.1994	2

Кстати, среди обработчиков есть “Слияние с другим узлом”, который позволяет объединить записи файлов справочников БИК.

### 2.3 Анализ данных по кредитам

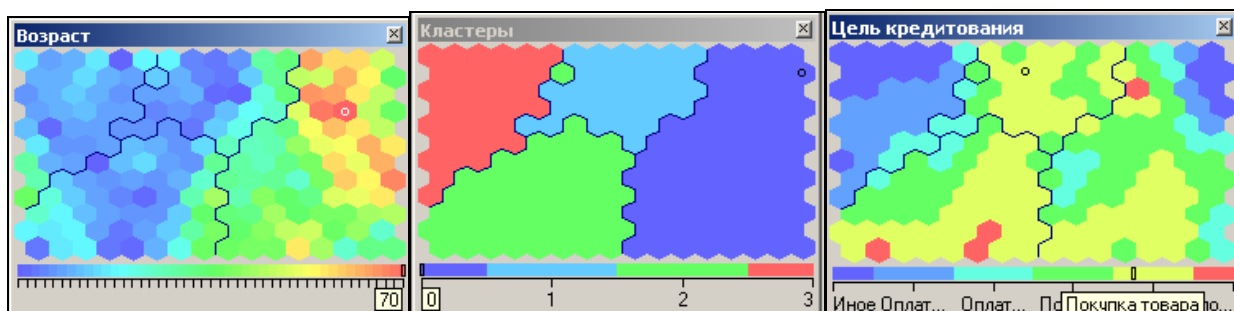
Загрузим в Deductor Academic файл Credit.txt из поставки программы, содержащий 149 записей о фактах выдачи виртуальной КО кредита заемщикам.

Построим обработчик “Карта Кохонена”. По двум картам “количество иждивенцев” и “давать кредит” видим, что, чем меньше иждивенцев, тем охотнее КО выдает кредит (признак выдачи кредита – число 1), т.к. максимумы (красные цвета) на картах асимметричны. При этом срок кредита и сумма кредита должны быть минимальны (синий цвет).



Для людей старшего возраста (выделена точка для значения “70 лет”) решения о выдаче кредитов отрицательные (признак выдачи кредита – число 0), т.к. в этой области карты “выдать кредит” присутствует синий цвет нулевого значения признака выдачи кредита.

Все множество записей хранилища данных разбито на 4 кластера заемщиков. Кластеру под №1 (голубого цвета), которому соответствуют отрицательные решения о выдаче кредита (синий цвет на кластере “давать кредит”), соответствуют большие суммы и сроки кредита, на диаграмме “цель кредитования” этой группы заемщиков в большинстве случаев указано “покупка товара”



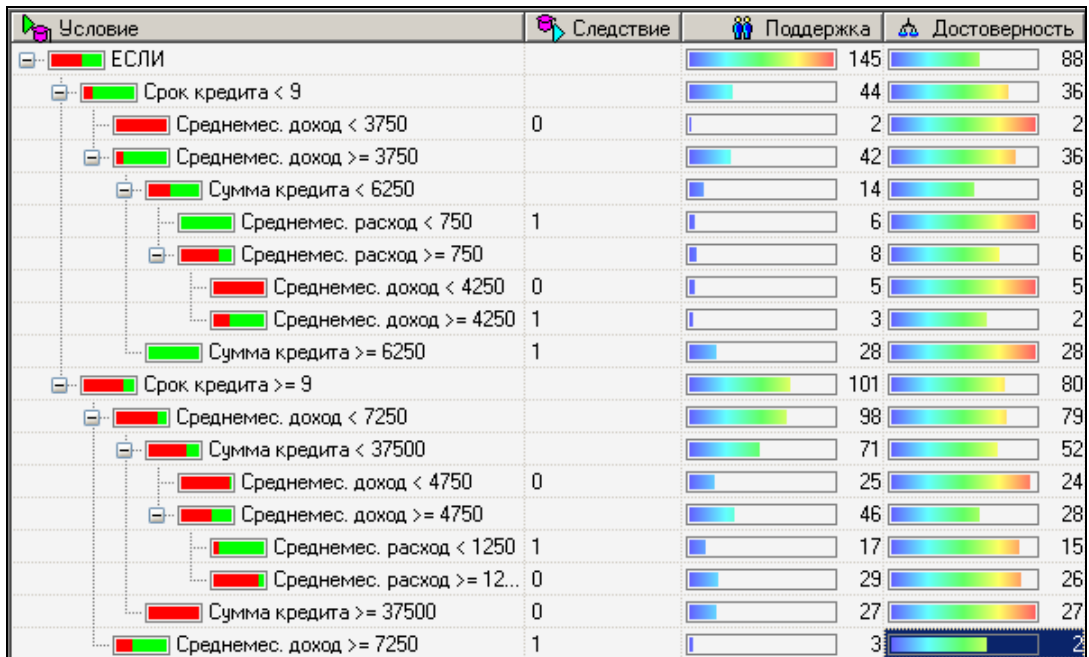
По таблице сопряженности можно видеть, что в 87-и случаях решение о выдаче кредита было отрицательным, в 49-и случаях решение было положительным и эти решения соответствуют модели, поддерживаемой программой. В 10-и случаях решение было положительным, однако кредит, в соответствии с моделью, не надо было выдавать, с другой стороны, в 3-х случаях отрицательного решения о выдаче кредита кредит можно было выдать

Фактически	Классифицировано		
	0	1	Итого
0	87	3	90
1	10	49	59
Итого	97	52	149

По отдельным случаям кредитов Deductor Academic уменьшает количество решений по выдаче кредитов (детализация записи выполняется при выборе элемента таблицы) или предлагает иное решение.

Модель дерева решений для оценки кредитоспособности заемщика позволяет понять решения КО: положительное решение принималось для случая краткосрочных кредитов - при малых средних расходах в бюджете семьи, или при больших доходах, или при больших суммах кредита. Для случая долгосрочных кредитов положительное решение принималось при малых расходах или при больших доходах.

Отрицательное решение принималось при низких доходах или больших расходах бюджета семьи



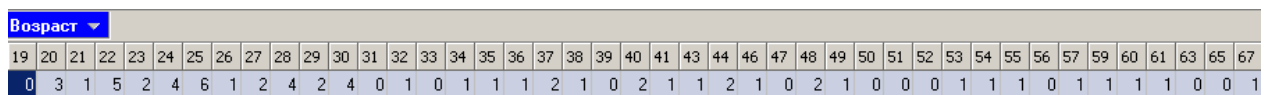
По визуализатору “значимость атрибутов” понятно влияние исходных данных на решение о выдаче кредита – срок кредита, расход и доход семьи, а также сумма кредита являлись основными факторами при принятии решения

Дерево решений X    Значимость атрибутов X

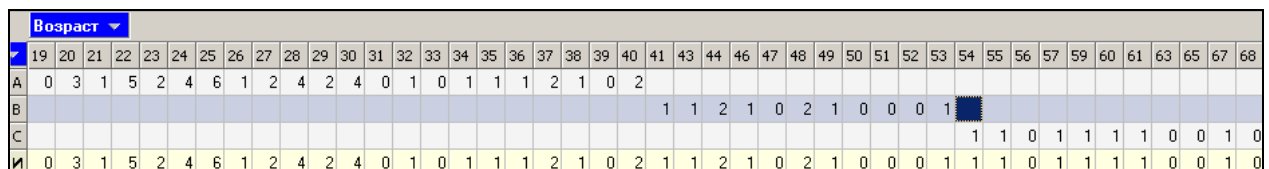
Целевой атрибут: Давать кредит(число)

№	Номер	Атрибут	Значимость, %
1	2	Срок кредита	33,641
2	6	Среднемес. расход	27,334
3	1	Сумма кредита	19,931
4	5	Среднемес. доход	19,094
5	7	Количество иждивенцев	0,000
6	3	Цель кредитования	0,000
7	4	Возраст	0,000

Зависимость решения о выдаче кредита от возраста показывает, что чаще положительные решения были приняты для возрастного интервала 25-30 лет



Использование обработчика “калькулятор” позволяет провести ABC-анализ, согласно которому 75% людей, обращающихся за кредитами относится к возрастной группе от 19 до 40 лет



Из-за отсутствия в исходном файле истории выплат кредитов, невозможно выполнить проверку надежности заемщика. Смоделируем эту ситуацию в соответствии с нашей собственной логикой, введя в хранилище данных дополнительное поле “просрочка” с помощью обработчика “калькулятор” по формуле

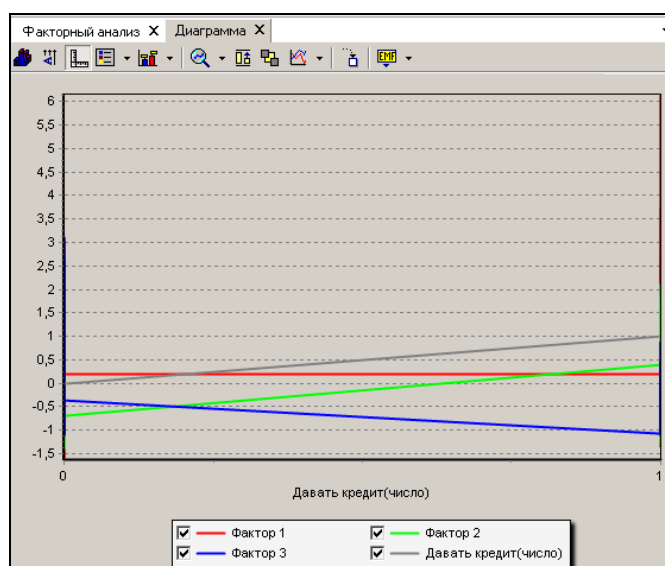
$iff((\text{возраст} > 50 \text{ and } \text{кол-во иждивенцев} \geq 3) \text{ or } (\text{доход} < 5500 \text{ and } \text{расход} \geq 1500)); 1; 0)$

Согласно формуле заемщик в возрасте 50 лет и старше с количеством иждивенцев больше или равном 3, или расходом, достигающем 35% дохода будет отнесен к разряду неблагонадежных со значением 1 в дополнительном поле “просрочка”.

Логика поддержана результатами факторного анализа, показывающего значимость входных данных,

Переменные	Итоговые факторы (Варимакс)		
	Фактор 1	Фактор 2	Фактор 3
Среднемес. доход	0,9633		
Среднемес. расход	0,9572		
Сумма кредита			0,9508
Срок кредита			0,9863
Количество иждивенцев		0,9276	
Возраст		0,9304	

и одинаковым трендом линий “давать кредит” и “Фактор 2: количество иждивенцев и возраст”



Куб, построенный на измененных данных, показывает число

ошибочных решений о выдаче кредита, и оно равно 8.

Дата кредитования		Цель кредитования	
Давать кредит			
просрочка	Да	Нет	Итого:
0	51	0	51
1	8	0	8
<b>Итого:</b>	59	0	59

Выполним обработку “квантование” для деления множества данных на возрастные интервалы до 26 лет, от 26 до 39 лет, после 39 лет. На результирующем множестве выполним обработку “логистическая регрессия”.

По таблице сопряженности видно, что в трех случаях отсутствия просрочки система выдала решение о ее наличии. В одном случае просрочки не было определено ее наличие. Большая часть случаев моделью классифицирована правильно, а именно 36 просрочек и 109 случаев отсутствия просрочек смоделировано и определено

Таблица сопряженности		Классифицировано		Итого
Фактически	0	1		
0	109	3		112
1	1	36		37
<b>Итого</b>	110	39		149

Наконец, по возрастным категориям признак выдачи кредита распределен следующим образом

Возраст	Σ Давать кредит(число)
до 26	21
от 26 до 39	20
от 39	18
<b>Итого:</b>	59

В работе [2] в анализируемой выборке данных присутствует признак наличия задолженности. Воспользуемся выводами, сделанными на основе обработки их данных. На основе вывода “Выдавать кредит не стоит следующим лицам:

– заемщикам, не состоящим в браке, большая часть которых не имеет собственного жилья;

– заемщикам в возрасте от 39 лет, состоящим в браке и имеющим иждивенцев

построим обработку “калькулятор”, введя в исходный набор данных столбец “просточка2” по формуле

$iff((COL32="Нет" \text{ and } COL13="нет жилья в собственности") \text{ or } (COL7>39 \text{ and } COL32="Да" \text{ and } COL31>0));1;0)$

Число ошибочных решений о выдаче кредита равно 14

Дата кредитования		Цель кредитования		
		Давать кредит		
просрочка2	Да	Нет	Итого:	
0	45	0	45	
1	14	0	14	
<b>Итого:</b>	59	0	59	

Вновь выполним квантование и логистическую регрессию.

По таблице сопряженности видно (в скобках представлен первый результат для сравнения), что в одном случае отсутствия просрочки система выдала решение о ее возможном наличии. В одном случае просрочки не было определено ее наличие. Большая часть случаев моделью классифицирована правильно, а именно 41 просрочка и 106 случаев отсутствия просрочек и смоделировано и определено

Таблица сопряженности X		Кoeffициенты ре		
просрочка2				
		Классифицировано		
Фактически	0	1	Итого	
0	106	1	107	
1	1	41	42	
<b>Итого</b>	107	42	149	

Таблица сопряженности X		Кoeffициенты ре		
просрочка				
		Классифицировано		
Фактически	0	1	Итого	
0	109	3	112	
1	1	36	37	
<b>Итого</b>	110	39	149	

Результаты показывают, что наши предположения об алгоритме выдачи кредитов соответствуют реальной политике КО, несмотря на то, что исходные условия моделирования отличаются.

### Заключение.

Научной новизной исследований явилось применение известного

метода анализа данных для получения не только итогов, но и тенденций в исследуемых данных.

Проделанная работа актуальна, т.к. мы расширили область применения известной технологии и изменили точку зрения на ее возможности. Эта задача, обычно, выполняется дорогими системами поддержки принятия решений, для которых необходимо также сопровождение и обучение работе с ними.

### **Литература**

1. Аналитическая платформа Deductor [Электронный ресурс]. — Режим доступа. — URL: <http://www.basegroup.ru>
2. Раменская Алина Владимировна. Моделирование кредиторской задолженности заемщиков коммерческого банка. Тезисы конференции, Оренбургский государственный университет.
3. Электронный курс Data Mining. — Режим доступа. — URL: <http://www.kodges.ru/komp/132934-data-mining.html>