# Введение

Интеллектуальный анализ данных, или Data Mining, – это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Современные технологии *Data* *Mining* (*discovery-driven data mining*) обрабатывают информацию с целью автоматического поиска шаблонов (паттернов), характерных для каких-либо фрагментов неоднородных многомерных данных. В отличие от оперативной аналитической обработки данных (*online analytical processing, OLAP*) в *Data Mining* бремя формулировки гипотез и выявления необычных шаблонов переложено с человека на компьютер.

Выделяют пять стандартных типов закономерностей (задач), которые позволяют выявлять методы *Data Mining*: ассоциация, последовательность, классификация, кластеризация, прогнозирование.

Закономерность типа а*ссоциация* наблюдается в данных, когда несколько событий связаны друг с другом и происходят при этом одновременно. Например, исследование, проведенное в супермаркете, может показать, что 65% купивших кукурузные чипсы берут также и «кока-колу», а при наличии скидки за такой комплект «колу» приобретают в 85% случаев. Располагая сведениями о подобной ассоциации, менеджерам легко оценить, насколько действенна предоставляемая скидка.

Закономерность типа «*последовательность»* предполагает наличие в данных цепочки связанных друг с другом и распределенных во времени событий. Так, например, после покупки дома в 45% случаев в течение месяца приобретается и новая кухонная плита, а в пределах двух недель 60% новоселов обзаводятся холодильником.

Закономерность типа «*классификация»* выявляется в данных на основе анализа признаков уже классифицированных объектов, при этом известна принадлежность объектов к классам. Результатом является формирование правил отнесения объектов к классам.

Закономерность типа «*кластеризация»* предполагает наличие в данных сходых по каким-либо признакам групп объектов, причем количество групп и принадлежность объектов к ним заранее не заданы. С помощью кластеризации средства *Data Mining* самостоятельно выделяют различные однородные группы данных.

Поиск закономерности типа «*прогнозирование»* проводится на основе информации, хранящейся в базах данных в виде временных рядов. Если удается построить математическую модель и найти шаблоны, адекватно отражающие эту динамику, есть вероятность, что с их помощью можно предсказать и поведение системы в будущем.

# Раздел 1Описание объекта исследования, как системы

Системный подход реализуется в методологии системного анализа, который является одним из эффективных инструментов исследования социально-экономических и политических процессов.

# 

# 1.1 Цель объекта исследования

Объектом исследования являются показатели деятельности кредитных организаций. Целью объекта исследования является отражение динамики процесса взаимодействия населения и кредитных организаций.

# 

# 1.2 Структура системы, описание внешней среды

Тип описываемой структуры – плоская (горизонтальная декомпозиция) (схема1)

Схема 1

Ипотечный

Потребительский

Для лизинга

Уровень доверия между населением и банками

*Кредит*

НТП

Индивидуальным предпринимателем

Банковская ссуда

На уровне двух стран

**Ставка** рефинансирования

**Доход Населения**

# 1.3 Классификация системы

Данная система является *открытой*, т.к. основные элементы системы – кредитор и заемщик – обеспечивают постоянное циркулирование денежных средств с дачей денег в долг (входом) денежных средств со стороны кредитора и выплат (выхода) полученной суммы плюс проценты со стороны заемщик.

Описываемая система – *гетерогенная*. Доказательством данного факта является многообразие видов кредита, которые изображены на схеме 1.

Система является *управляемой*. В роли управленцев этой системы выступают кредитор и заемщик. А так как они являются в тоже время и элементами данной системы, то данную систему можно считать самоуправляемой.

# Раздел 2 Интеллектуальный анализ данных о закономерностях

# функционирования объекта исследования

Цель анализа состоит в применении интеллектуального анализа данных для исследования заданной структуры данных для разработки принятия решений.

В данной работе основной задачей является выявление различий между регионами России на основе показателей деятельности кредитных организаций приведенным на 01.0.1.2007.

**Исходные данные**

**Объектами** в данной работе являются: Центральный, Северо-Западный, Южный, Приволжский, Уральский, Сибирский, Дальневосточный федеральный округа.

В качестве **признаков** в данной работе выступают: **объем выданных кредитов физическим лицам, объем кредитов физическим лицам на покупку жилья, средневзвешенный срок кредитования, средневзвешенная процентная ставка, объем ипотечных жилищных кредитов физическим лицам, средневзвешенный срок кредитования (по ипотечным кредитам), средневзвешенная процентная ставка (по ипотечным кредитам).**

# 2.1 Анализ методом главных компонент

Метод главных компонент является методом визуализации данных.

Методы визуализации данных нацелены на поиск наиболее выразительных изображений совокупности исследуемых объектов для последующего максимального использования потенциала зрительного анализатора экспериментатора.

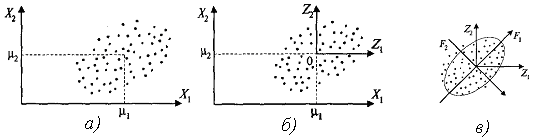
Визуализация данных предполагает получение тем или иным способом графического отображения совокупности объектов в новое координатное пространство, в качестве которого могут быть использованы либо числовая ось, либо плоскость, либо трехмерное пространство, максимально отражающие особенности распределения этих объектов в многомерном пространстве.

В соответствии с количеством измерений в новом координатном пространстве могут быть следующие способы визуального представления:

* одномерное измерение, или *1-D*,
* двумерное измерение, или *2-D*,
* трехмерное (проекционное) измерение, или *3-D*.

С геометрических позиций алгоритм построения главных компонент (ГК) состоит в следующем.

Производится центрирование исходных данных (рис.1а), начало координат переносится в центр распределения данных (центроид), являющимся центром эллипсоида рассеивания случайного вектора *Х* (рис.1б).



**Рис. 2. Формирование главных компонент**

Затем осуществляется формирование главных компонент *F1, F2,…,Fp* (рис.1в). Линейные комбинации выбираются таким образом, что среди всех возможных комбинаций первая главная компонента *F1(X)* обладает наибольшей дисперсией. Дисперсия σ стремится к максимуму: *F1(X)=I=max* σ ->*ωpi*. Графически это выглядит как ориентация новой координатной оси *F1* вдоль направления наибольшей вытянутости эллипсоида рассеивания объектов в исходном пространстве *P* признаков. а разбросом вдоль нескольких последних осей можно пренебречь.

Вторая главная компонента *F2(X)* перпендикулярна первой и строится исходя из предположений нахождения максимальной дисперсии среди всех оставшихся линейных комбинаций, некоррелированных с первой ГК.

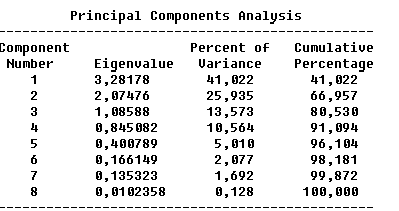
Остальные главные компоненты определяются аналогичным способом.

Количество главных компонент равно количеству элементарных признаков.

Введем исходные данные в электронную таблицу STATGRAPHICS (48\*8).

Анализ методом главных компонент состоит из нескольких этапов:

1. Необходимо получить данные о результатах анализа (таблица сводки) (рис.3)

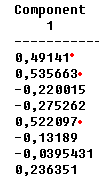


**рис.3 Сводка метода ГК**

В Данной таблице содержится информация о результатах построения главных компонент: собственные значения главных компонент (*eigenvalue*), упорядоченных по величине, процент дисперсии (*percent of variance*), приходящейся на каждую выделенную главную компоненту, накопленный процент дисперсии (*cumulative percentage*).

На основе полученных результатов можно седлать вывод о том, что первые 2 компоненты дают 80 процентов дисперсии, что достаточно для проведения анализа распределения компонентов в пространстве.

1. Далее проанализируем таблицу весов признаков (рис.4)



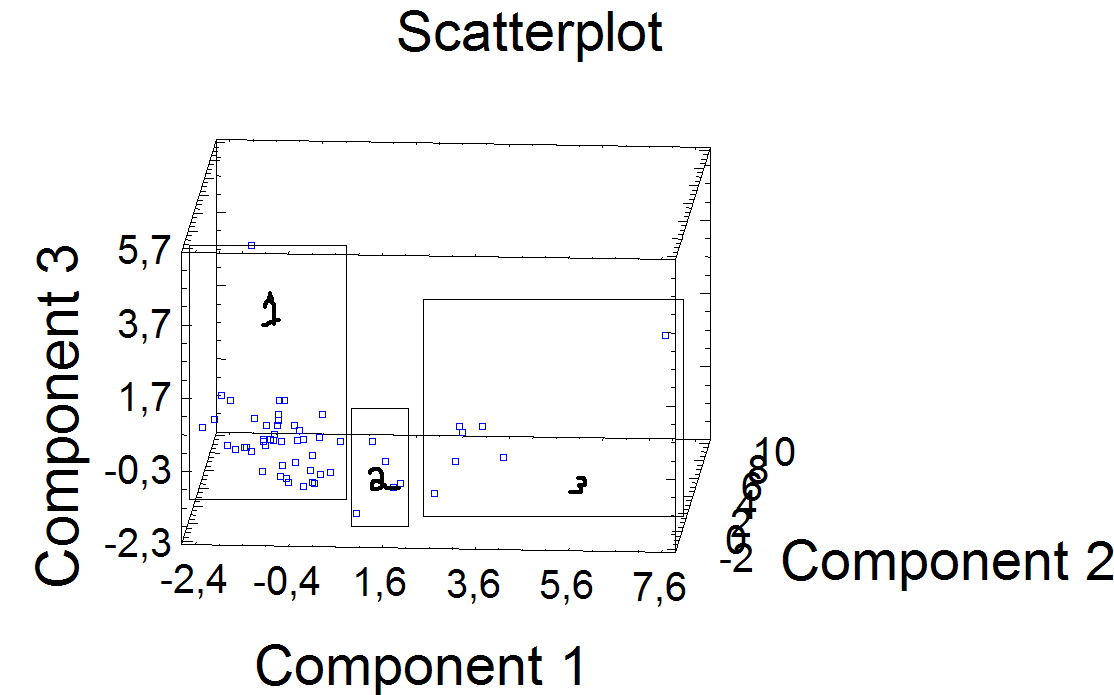
**рис.4 Веса признаков в главных компонентах**

Определим зависимость первой главной компоненты от признаков. Зависимость от того или иного признака определяется величиной признака.

Первая главная компонента в наибольшей степени зависит от 3 признаков:

**объем кредитов, выданных физическим лицам, объем кредитов, выданных физическим лицам на покупку жилья, объем выданных ипотечных жилищных кредитов физическим лицам.**

Ниже представлена проекция исследуемого множества на пространство трех ГК (рис.5).



**рис.5 Проекция исследуемых объектов в пространство трех ГК**

Из рис.5 видно, что исследуемая совокупность разделилась на три класса.

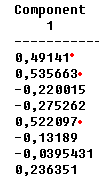
3) Определим значащие признаки в составе главных компонент по следующей формуле:

, где [*wkj*] – подмножество, участвующих в названии весовых коэффициентов *j-*й компоненты,



[*wj*] – все весовые коэффициенты j-й компоненты.

Ниже представлена таблица весовых коэффициентов первой ГК.



В соответствии с вышепредставленной формулой коэффициент информативности рассчитывается следующим образом:

= 0,8



Ки1 принадлежит интервалу [0,75,0,95], что говорит об определении ГК1 влиянием следующих признаков: *объем выданных кредитов физическим лицам, объем кредитов, выданных физическим лицам на покупку жилья, объем выданных ипотечных жилищных кредитов физическим лицам.*

4) Правило классификации на основе анализа методом главных компонент выглядит следующим образом:

**ГК1**=*объем выданных кредитов физическим лицам* **\*0,49+** *объем кредитов, выданных физическим лицам на покупку жилья* **\*0,53+** *объем выданных ипотечных жилищных кредитов физическим лицам* **\*0,****52**

**Если** *объем выданных кредитов физическим лицам = малое, объем кредитов, выданных физическим лицам на покупку жилья = малое, объем выданных ипотечных жилищных кредитов физическим лицам =малое, объем выданных кредитов индивидуальным предпринимателям = малое, то класс = 1*

**Если** *объем выданных кредитов физическим лицам = среднее, объем кредитов, выданных физическим лицам на покупку жилья = среднее, объем выданных ипотечных жилищных кредитов физическим лицам = среднее, объем выданных кредитов индивидуальным предпринимателям = среднее, то класс = 2*

**Если** *объем выданных кредитов физическим лицам = большое, объем кредитов, выданных физическим лицам на покупку жилья = большое, объем выданных ипотечных жилищных кредитов физическим лицам = большое, объем выданных кредитов индивидуальным предпринимателям = большое, то класс = 3*

В **первый** класс вошли объекты под номерами: 1-9, 11-17,19-28, 30-37, 39-42, 48,50,53,56.

Во **второй** класс вошли объекты под номерами: 42, 47, 51,38,52,55,45

В **третий** класс вошел один объект под номером: 10,18,29,43,46,49,54

# 2.2 Кластерный анализ

Кластерный анализ предназначен для разбиения множества объектов на заранее неизвестное или в редких случаях заданное количество групп (кластеров) на основании некоторого математического критерия качества классификации кластеризации.

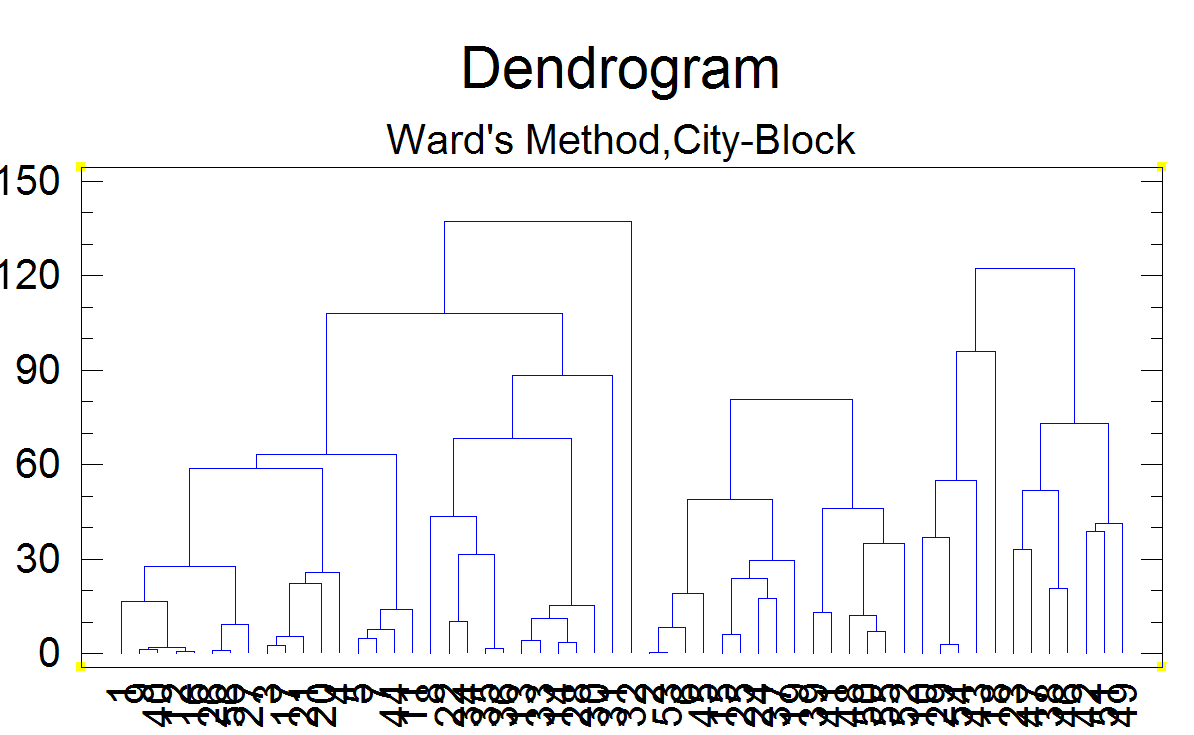
Введем исходные данные в электронную таблицу Statgraphics. Первый этап анализа – построение дендрограммы. Нажать на кнопку «*Graphical* *options*» меню окна «*Cluster* *Analysis*», выбрать отображение в виде дендрограммы (*Dendrogram*) и нажать ОК.

На дендрограмме видны три дерева. По вертикальной оси отложено расстояние для каждого шага работы агломеративного иерархического алгоритма кластеризации. На горизонтальной оси показаны наблюдения, скомбинированные в соответствии с проведенным анализом.

Поскольку на дендрограмме можно увидеть отчетливую картину трех группировок и имен наблюдений, вошедших в выделенные кластеры, то следовательно, для более подробного рассмотрения группировок следует задать их количество равным 3.

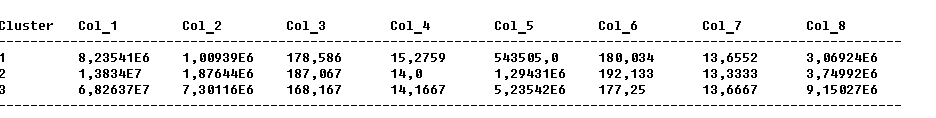
Введем исходные данные в электронную таблицу STATGRAPHICS (50\*8).

Первый шаг анализа - построение дендрограммы для трех кластеров (рис.6). При построении необходимо учитывать виды расстояний: между объектами – «City-Block», между кластерами «Ward’s Method».



**рис.6 Трехкластерная дендрограмма**

Второй шаг получение сводки кластерного анализа (рис.7).



**рис.7 Сводка кластерного анализа**

Данная сводка содержит информацию о числе кластеров, количестве объектов в каждом кластере и соответствующем проценте населенности.

По координатам центроидов можно судить о том, какие переменные играют наиболее важную роль в каждом кластере, а также составить правило классификации. В полученной сводке есть признаки, кластерные значения которых не составляют большую разницу между собой. К таким признакам относятся: **средневзвешенный срок кредитования(col\_3), средневзвешенная процентная ставка(col\_4), средневзвешенный срок кредитования для ипотечных кредитов(col\_6) и средневзвешенная процентная ставка для ипотечных кредитов(col\_7).**

Правило классификации для первого кластера:

**Если** объем выданных кредитов физическим лицам =большой, объем кредитов, выданных физическим лицам на покупку жилья = малое, объем выданных ипотечных жилищных кредитов физическим лицам = малое, объем выданных кредитов индивидуальным предпринимателям = малое, то класс = 1.

Правило классификации для второго кластера:

**Если** объем выданных кредитов физическим лицам = малый, объем кредитов, выданных физическим лицам на покупку жилья = среднее, объем выданных ипотечных жилищных кредитов физическим лицам = среднее, объем выданных кредитов индивидуальным предпринимателям = среднее, то класс = 2.

Правило классификации для третьего кластера:

**Если** объем выданных кредитов физическим лицам = средний, объем кредитов, выданных физическим лицам на покупку жилья = большой, объем выданных ипотечных жилищных кредитов физическим лицам =большой, объем выданных кредитов индивидуальным предпринимателям = большой, то класс = 3.

**2.3 Дерево решений**

*Деревья решений* – это способ представления правил в иерархической, последовательной структуре. В у*зле дерева* осуществляется проверка значения некоторой независимой переменной. Если переменная, которая проверяется в узле, принимает категориальные значения, то каждому возможному значению соответствует ветвь, выходящая из узла дерева. Если значением переменной является число, то проверяется, больше или меньше это значение некоторой константы.

*Листья деревьев* соответствуют классам. Каждый лист дерева представляет собой один класс. Путь, ведущий от корня дерева к этому листу, соответствует правилу классификации. Отметим, что один и тот же класс может быть указан в нескольких листьях дерева. Однако, каждому пути к отдельному листу (правилу классификации) соответствует множество объектов, в котором элементы (объекты) не могут повторяться в разных листьях.

Построение ДР проводится с использованием пакета See5, позволяющим конструировать классификатор объектов в виде ДР, которому может быть поставлено в соответствие некоторое множество логических правил.

Перед вычислением дерева необходимо скопировать данные в блокнот word pad, после чего изменить формат файла на *data*. Исходные данные для вычисления дерева решения примут следующий вид (имя файла derevo):

1,1826363,189,15

2,917535,188,14

1,1482830,166,15

1,2474846,164,15

1,642738,167,15

2,1106109,190,14

1,462245,147,15

1,1105639,197,16

1,788815,189,15

3,7853229,156,14

1,451150,160,15

1,1420937,189,15

1,1266100,203,15

1,438172,192,15

2,1682123,184,14

1,1263580,184,15

1,1644670,172,15

3,14438366,112,14

2,1229661,175,13

1,4356096,183,15

1,1851976,177,15

1,230977,185,16

3,2383630,170,14

2,1374942,160,14

2,1359242,192,14

1,758176,179,15

1,944696,182,15

1,438416,197,15

3,8669422,172,14

1,211300,198,15

1,122290,89,18

1,32493,134,15

1,577318,204,15

1,136223,186,16

1,209524,217,16

1,787277,209,16

2,36251,171,14

3,4884241,190,14

2,3403749,195,15

1,1013384,188,15

2,2853828,202,15

3,4082560,171,15

3,10950662,172,14

1,921844,154,15

2,940159,189,14

3,7889566,165,14

3,5052323,167,14

2,3311362,199,14

3,6933937,180,14

2,2734210,192,14

3,6301381,187,15

2,3518255,191,13

2,1015552,188,14

3,8174605,176,14

2,2663564,190,14

1,1412219,178,15

После этого создаем файл под именем *derevo* формат *names*.

*Файл derevo.names* выглядит следующим образом:

Class. |target attribute

Class: 1,2,3.

объем выданных кредитов физическим лицам*:* continuous.

средневзвешенный срок кредитования*:* continuous.

средневзвешенная процентная ставка*:* continuous.

Шаг 1.

Построение дерева решений.

Decision tree:

средневзвешенная процентная ставка <= 14:

:...объем выданных кредитов физ.лицам <= 4082560: 2 (14/1)

: объем выданных кредитов физ.лицам > 4082560: 3 (9)

средневзвешенная процентная ставка > 14:

:...объем выданных кредитов физ.лицам <= 2663564: 1 (28)

объем выданных кредитов физ.лицам > 2663564:

:...объем выданных кредитов физ.лицам <= 3518255: 2 (2)

объем выданных кредитов физ.лицам > 3518255: 3 (3/1)

В полученном дереве 5 ветвей. Первая ветвь: 2 класс, состоящий из 14-ти объектов, причем 1 классифицируется ошибочно.

Вторая ветвь: 3 класс, состоящий из 9-ти объектов. Третья ветвь: 1 класс – 28 объектов. Четвертая ветвь: 2 класс – 2 объекта. Пятая ветвь: 3 класс – 3 объекта, причем 1 объект классифицируется ошибочно.

Данное дерево решений содержит в себе следующую информацию:

**Если средневзвешенная процентная ставка <= 14 и объем выданных кредитов физ. лицам<= 4082560, то класс 2 (14/1 объектов),**

**Если средневзвешенная процентная ставка <= 14 и объем выданных кредитов физ. лицам >4082560, то класс 3 (9 объектов),**

**Если средневзвешенная процентная ставка > 14 и объем выданных кредитов физ. лицам <= 2663564, то класс 1 (28 объектов),**

**Если средневзвешенная процентная ставка > 14 и объем выданных кредитов физ. лицам на покупку жилья >2663564 и <=3518255 то класс 2 (2 объекта).**

**Если средневзвешенная процентная ставка > 14 и объем выданных кредитов физ. лицам на покупку жилья >2663564 и >3518255 то класс 3 (3/1 объекта).**

Ниже представлены извлеченные правила.

Extracted rules:

Rule 1: (28, lift 1.9)

объем выданных кредитов физ.лицам <= 2663564

средневзвешанная процентная ставка > 14

-> class 1 [0.967]

Rule 2: (14/1, lift 3.3)

объем выданных кредитов физ.лицам <= 4082560

средневзвешанная процентная ставка <= 14

-> class 2 [0.875]

Rule 3: (5/3, lift 1.6)

объем выданных кредитов физ.лицам > 2663564

средневзвешанная процентная ставка > 14

-> class 2 [0.429]

Rule 4: (12/1, lift 4.0)

объем выданных кредитов физ.лицам > 3518255

-> class 3 [0.857]

Каждое правило имеет следующую структуру:

* номер правила,
* в скобках – количество объектов обучающей выборки,
* запись условной части правила (часть «ЕСЛИ»),
* после знака импликации (->) – запись заключительной части правила (часть «ТО»), в которой указана принадлежность к классу,
* величина, принимающая значение от 0 до 1, которая выражает степень доверия к правилу.

Decision Tree Rules

---------------- ----------------

Size Errors No Errors

5 2( 3.6%) 4 2( 3.6%) <<

(a) (b) (c) <-classified as

---- ---- ----

28 1 (a): class 1

15 (b): class 2

1 11 (c): class 3

1. Из 1 класса правильно классифицируются 28 объектов, 1 объект ошибочно относится к классу 3
2. Из 2 класса все 15 объектов классифицируются верно,
3. Из класса 3 верно классифицируются 11 объектов, один объект ошибочно относится к классу 2.

Количество сработавших правил равно 4, имеет место ошибка извлечения объектов (3.6 процента).

# Раздел 3 Проблемы управления функционированием объекта

# исследования

С момента появления кредитной услуги единственным органом, предоставляющим таковую, являлся банк. Взяв необходимую сумму, заемщик осуществлял сделки по покупке, использованию различных услуг, взятию определенного имущества в аренду и т.д. Однако данная структура системы являлась недостаточно эффективной по причине больших нагрузок на деятельность банков. Данная проблема была решена путем применения **принципа декомпозиции** (*возможность расчленения систем по тому или иному признаку на отдельные части (подсистем) и в формировании для них собственных целей и функций, исходя из условия обеспечения достижения глобальных целей системы*). Представленная в данной работе схема 1 отображает разновидность предоставляемых кредитных услуг в зависимости от целей потребителей.

Активность кредитных организаций определяется объемом выданных кредитов гражданам. Характер тенденции в заключении сделок зависит от степени доверия между заемщиком и кредитором. Однако практика российского кредитования указала на отрицательную тенденцию в направлении доверия между банками и населением. Причиной этому послужило отсутствие предоставление данных со стороны банка о полной сумме процента с взятой суммы, вернуть которую в определенные сроки должен заемщик. Вследствие этого заемщик со своей стороны очень часто оказывается не в состоянии оплатить кредит в указанные сроки.

Данную проблему можно отнести к разновидности несоблюдения **принципа управляемости и наблюдаемости** *(управляемость предполагает воздействие на элементы системы в процессе управления, а наблюдаемость предполагает возможность получения информации о состоянии системы в процессе управления).*

В данном случае несоответствие составляющей принципа «наблюдаемость», выражено в отсутствии информации о полной сумме кредита. Заемщик в свою очередь не мог воздействовать на процесс выдачи информации о полной сумме.

Однако принятие закона, предписывающего выдавать полную информацию обо всей сумме кредита, положило начало пути по решению данной проблемы.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Номер | Регион | Кластер | Компонентный анализ | Дерево решений |
| 1 | Белгородская область | 1 | 1 | 1 |
| 2 | Брянская область | 2 | 1 | 2 |
| 3 | Владимирская область | 1 | 1 | 1 |
| 4 | Воронежская область | 1 | 1 | 1 |
| 5 | Ивановская область | 1 | 1 | 1 |
| 6 | Калужская область | 2 | 1 | 2 |
| 7 | Костромская область | 1 | 1 | 1 |
| 8 | Курская область | 1 | 1 | 1 |
| 9 | Липецкая область | 1 | 1 | 1 |
| 10 | Московская область | 3 | 3 | 3 |
| 11 | Орловская область | 1 | 1 | 1 |
| 12 | Рязанская область | 1 | 1 | 1 |
| 13 | Смоленская область | 1 | 1 | 1 |
| 14 | Тамбовская область | 1 | 1 | 1 |
| 15 | Тверская область | 2 | 1 | 2 |
| 16 | Тульская область | 1 | 1 | 1 |
| 17 | Ярославская область | 1 | 1 | 1 |
| 18 | г. Москва | 3 | 3 | 3 |
| 19 | Республика Карелия | 2 | 1 | 2 |
| 20 | Республика Коми | 1 | 1 | 3 |
| 21 | Архангельская область | 1 | 1 | 1 |
| 22 | Ненецкий автономный округ | 1 | 1 | 1 |
| 23 | Вологодская область | 3 | 1 | 2 |
| 24 | Калининградская область | 2 | 1 | 2 |
| 25 | Ленинградская область | 2 | 1 | 2 |
| 26 | Мурманская область | 1 | 1 | 1 |
| 27 | Новгородская область | 1 | 1 | 1 |
| 28 | Псковская область | 1 | 1 | 1 |
| 29 | г. Санкт-Петербург | 3 | 3 | 3 |
| 30 | Республика Адыгея (Адыгея) | 1 | 1 | 1 |
| 31 | Республика Дагестан | 1 | 1 | 1 |
| 32 | Республика Ингушетия | 1 | 1 | 1 |
| 33 | Кабардино-Балкарская Республика | 1 | 1 | 1 |
| 34 | Республика Калмыкия | 1 | 1 | 1 |
| 35 | Карачаево-Черкесская Республика | 1 | 1 | 1 |
| 36 | Республика Северная Осетия-Алания | 1 | 1 | 1 |
| 37 | Чеченская Республика | 2 | 1 | 2 |
| 38 | Краснодарский край | 3 | 2 | 3 |
| 39 | Ставропольский край | 2 | 2 | 3 |
| 40 | Астраханская область | 1 | 1 | 1 |
| 41 | Волгоградская область | 2 | 1 | 3 |
| 42 | Ростовская область | 3 | 2 | 3 |
| 43 | Республика Башкортостан | 3 | 3 | 3 |
| 44 | Республика Марий Эл | 1 | 1 | 1 |
| 45 | Республика Мордовия | 2 | 2 | 2 |
| 46 | Республика Татарстан (Татарстан) | 3 | 3 | 3 |
| 47 | Удмуртская Республика | 3 | 2 | 3 |
| 48 | Чувашская Республика - Чувашия | 2 | 1 | 2 |
|  | Пермский край |  |  |  |
| 49 | Кировская область | 3 | 3 | 3 |
| 50 | Нижегородская область | 2 | 1 | 3 |
| 51 | Оренбургская область | 3 | 2 | 3 |
| 52 | Пензенская область | 2 | 2 | 2 |
| 53 | Самарская область | 2 | 1 | 2 |
| 54 | Саратовская область | 3 | 3 | 3 |
| 55 | Ульяновская область | 2 | 2 | 3 |
| 56 | Курганская область | 1 | 1 | 1 |

**Заключение**

Правило классификации на основе метода главных компонент:

**Если** *объем выданных кредитов физическим лицам = малое, объем кредитов, выданных физическим лицам на покупку жилья = малое, объем выданных ипотечных жилищных кредитов физическим лицам =малое, объем выданных кредитов индивидуальным предпринимателям = малое, то класс = 1*

**Если** *объем выданных кредитов физическим лицам = среднее, объем кредитов, выданных физическим лицам на покупку жилья = среднее, объем выданных ипотечных жилищных кредитов физическим лицам = среднее, объем выданных кредитов индивидуальным предпринимателям = среднее, то класс = 2*

**Если** *объем выданных кредитов физическим лицам = большое, объем кредитов, выданных физическим лицам на покупку жилья = большое, объем выданных ипотечных жилищных кредитов физическим лицам = большое, объем выданных кредитов индивидуальным предпринимателям = большое, то класс = 3*

Правило классификации на основе кластерного анализа:

Правило классификации для первого кластера:

**Если** объем выданных кредитов физическим лицам =большой, объем кредитов, выданных физическим лицам на покупку жилья = малое, объем выданных ипотечных жилищных кредитов физическим лицам = малое, объем выданных кредитов индивидуальным предпринимателям = малое, то класс = 1.

Правило классификации для второго кластера:

**Если** объем выданных кредитов физическим лицам = малый, объем кредитов, выданных физическим лицам на покупку жилья = среднее, объем выданных ипотечных жилищных кредитов физическим лицам = среднее, объем выданных кредитов индивидуальным предпринимателям = среднее, то класс = 2.

Правило классификации для третьего кластера:

**Если** объем выданных кредитов физическим лицам = средний, объем кредитов, выданных физическим лицам на покупку жилья = большой, объем выданных ипотечных жилищных кредитов физическим лицам =большой, объем выданных кредитов индивидуальным предпринимателям = большой, то класс = 3.

На основе результатов дерева решений:

**Если средневзвешенная процентная ставка <= 14 и объем выданных кредитов физ. лицам<= 4082560, то класс 2 (14/1 объектов),**

**Если средневзвешенная процентная ставка <= 14 и объем выданных кредитов физ. лицам >4082560, то класс 3 (9 объектов),**

**Если средневзвешенная процентная ставка > 14 и объем выданных кредитов физ. лицам <= 2663564, то класс 1 (28 объектов),**

**Если средневзвешенная процентная ставка > 14 и объем выданных кредитов физ. лицам на покупку жилья >2663564 и <=3518255 то класс 2 (2 объекта).**

**Если средневзвешенная процентная ставка > 14 и объем выданных кредитов физ. лицам на покупку жилья >2663564 и >3518255 то класс 3 (3/1 объекта).**

На основе результатов интеллектуального анализа данных можно сделать следующий вывод.

Принадлежность региона к первому классу указывает на малый объем сделок между населением и кредитными организациями. К основной причине сложившейся ситуации можно низкий уровень доходности населения и высокие процентные ставки, вследствие чего имеют место многократные случаи невозвращения процентов по займу. К данному классу относятся регионы *Южного федерального округа*, несколько областей *Центрального Федерального округа.*

Классификация региона по второму классу говорит о средних показателях активности населения в сфере кредитования. Данный фактор свидетельствует о среднем уровне доходов населения, что в свою очередь говорит о возможности приобретать жилье и вести бизнес за счет заемных средств у банка. К таким регионам относятся Чеченская республика, Ставропольский край, Республика Мордовия и т.д.

Принадлежность региона к третьему классу говорит об успешности во взаимодействии населения кредитных услуг. Однако к таким регионам относятся малая доля регионов страны: Московская область, Республика Башкортостан, Республика Татарстан, Ростовская область. Данный факт говорит, как об относительно высокой доходности населения, так и об установившемся уровне доверия между заемщиками и кредиторами. Население этих регионов сознательно вступает во взаимодействие с кредитной системой для ведения бизнеса покупки товаров и жилья.