

# **ИСПОЛЬЗОВАНИЕ КЛАСТЕРНЫХ ПРОЦЕДУР ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ УПРАВЛЕНИЯ И ОБРАБОТКИ ИНФОРМАЦИИ В СИСТЕМАХ МАССОВОГО ОБСЛУЖИВАНИЯ**

**С.А. Королев;**

**Д.Ю. Минкин, доктор технических наук, профессор.**

**Санкт-Петербургский университет ГПС МЧС России.**

**Н.А. Орлов.**

**Российский государственный педагогический университет**

**им. А.И. Герцена**

Рассмотрены возможности повышения эффективности управления и обработки информации в системах массового обслуживания с помощью априорной типизации клиентов. Эта задача решена на примере сети кафе с использованием процедур кластерного анализа. Приведены и обоснованы основные значимые характеристики для анализа посетителей кафе, рассмотрены общие примеры маркетинговых воздействий на них. Отмечена возможность использования данного метода и для других задач.

*Ключевые слова:* кластерный анализ, маркетинговые решения, эффективность продаж, сегментация данных, системы массового обслуживания

## **CLUSTER PROCEDURES USE FOR ENHANCEMENT OF THE MANAGEMENT AND PROCESSING OF INFORMATION IN QUEUING SYSTEMS**

S.A. Korolev; D.Yu. Minkin.

Saint-Petersburg university of State fire service of EMERCOM of Russia.

N.A. Orlov. Herzen state pedagogical university of Russia

Discussed the possibility of increasing the efficiency of the management and processing of information in queuing systems with a priori typing of clients. This problem is solved using cluster analysis procedures on the example of cafe net. The authors of the article reviewed the basic and most significant characteristics of visitors to cafes, and considered common examples of marketing influences on them. Also the possibility of using this method for other tasks was mentioned.

*Keywords:* cluster analysis, marketing solutions, sales efficiency, data segmentation, queuing systems

Эффективность различного рода маркетинговых решений в значительной степени зависит от социальных, психологических и возрастных особенностей клиентской аудитории каждой конкретной коммерческой организации. К таким особенностям относятся платежеспособность клиента, его заинтересованность в тех или иных группах товаров, ориентация на выбор товара/услуги (себе или ребенку), реакция на различные методы (реклама и т.д.).

Выбор тех или иных маркетинговых действий требует глубокого понимания характеристик совокупности клиентов, которое возможно достичь путем проведения их статистических исследований. В процессе этих работ крайне важно выбрать вектор

параметров, описывающих каждого клиента, провести накопление данных, значимое по выборке клиентов и осуществить процедуры, позволяющие систематизировать эти данные в интересах последующих маркетинговых решений.

Рассмотрим один из таких возможных подходов на примере сети кафе, клиентами которых являются самые разные люди, что позволяет быстро и эффективно накопить требуемый статистический материал.

К числу характеристик клиентов, значимых для данного вида бизнеса на основе экспертных оценок, были отнесены следующие:

- гендерная характеристика (пол) – одна из базисных характеристик, обычно используемая в паре с характеристикой возраста, которая позволяет получить представление о размерах половозрастных групп клиентов;

- возраст/возрастная группа – значимая характеристика, дающая информацию о том, какие каналы коммуникации, формы и методы рекламы использовать для взаимодействия с клиентом. Принятые сокращения: У – учащийся, М – молодежь, В – взрослые, П – пожилые;

- средний чек – характеристика, определяющая «вес» клиентов выделенной группы в доходах компании и позволяющая формировать ценовую политику;

- употребление клиентом алкоголя используется для формирования предположений о сопутствующих товарах/блюдах, в которых будет заинтересован клиент, употребляющий/не употребляющий алкоголь.

- курит ли клиент, данная характеристика позволяет выделить процент курящих в кафе среди всего потока клиентов конкретного заведения и принять решение об изменении количества посадочных мест в зале для курящих и некурящих. Стоит отметить, что она в значительной степени потеряет свою значимость для исследования со вступлением в силу нового закона РФ, запрещающего курение в общественных местах, в том числе и в кафе, барах, ресторанах.

- пришел ли клиент с детьми, используется для исследования с целью выявления значимости наличия детского меню и иных удобств для конкретного заведения общественного питания.

В результате сбора данных о клиентах по вышеуказанным характеристикам, формируется сводная таблица первичных данных (табл. 1).

Таблица 1. Пример сводной таблицы первичных данных о клиентах кафе

№ клиента	Пол	Возраст	Возраст. группа	Средний чек, руб.	Алкоголь	Курение	С детьми?
1	М	21	У	400	НЕ ПЬЕТ	НЕ КУРИТ	НЕТ
2	М	25	М	550	НЕ ПЬЕТ	КУРИТ	НЕТ
3	Ж	36	В	330	ПЬЕТ	НЕ КУРИТ	ДА
4	Ж	19	У	250	НЕ ПЬЕТ	НЕ КУРИТ	НЕТ
5	Ж	22	М	285	НЕ ПЬЕТ	КУРИТ	НЕТ
...	...	...	...	...	...	...	...
N	М/Ж	14-77	У, М, В, П	120–1500	Н/П	Н/К	ДА/НЕТ

Очевидно, что вектора описываемых параметров могут быть больше, однако ограничимся шестью группами. При прочих равных условиях, чем больше значимых характеристик используется при сборе первичных данных для кластерного анализа, тем более конкретные рекомендации и более детерминированные управленческие решения могут предложить специалисты отдела маркетинга по результатам выполненной сегментации данных.

Высококвалифицированные специалисты в области маркетинга утверждают, что не существует общих рекомендаций для владельцев бизнеса общественного питания. К примеру, каждое кафе является уникальной организацией, так как имеет уникальные:

- бизнес-модель, миссию;
- генеральную, конкурентную и функциональные стратегии;
- план производства;
- организационный план (также включающий требования к отбору сотрудников, должностные инструкции для всех должностей и др.);
- позицию в конкурентной среде;
- а также такие характеристики как месторасположение кафе, его интерьер, структура меню, наличие покупательской преданности (лояльности потребителей) [1].

Тем не менее, зная социально-психологические портреты потребителей, можно в значительной степени влиять на эффективность продаж. Приведем примеры решений, основанные на таких априорных знаниях:

1. Использование различных каналов коммуникации для соответствующих половозрастных групп. В зависимости от возраста и пола, сознание потребителя по-разному воспринимает информацию, которую предоставляет ему внешняя среда [2].

Как правило, юноши и девушки в возрасте от 16 до 22 лет предпочитают взаимодействовать с продавцом товаров и услуг через мобильные средства связи, использующие доступ в интернет, так как используют мобильные устройства и компьютеры в повседневной жизни больше и чаще, чем все другие половозрастные группы вместе взятые. Можно сделать вывод о том, что наиболее действенной формой рекламы для этой группы потребителей (при прочих равных условиях) будет реклама в социальных сетях, поисковых системах (оптимизация официального сайта компании по ключевым запросам, которые используют потребители для поиска соответствующих товаров и услуг), социальных мобильных приложениях (к примеру, Foursquare). Также использование таких популярных инструментов мобильного маркетинга, как мобильные версии сайтов, электронные версии меню в свободном доступе, оправдано именно для этой группы потребителей, так как имеет наименьший порог вхождения для них.

Группа потребителей более старшего возраста, женщины и мужчины в возрасте от 50 лет, с большей охотой взаимодействует, к примеру, с официантами и бумажной версией меню, так как порог вхождения использования электронных средств связи для заказа блюд для них может быть слишком высок, а формирующийся ассоциативный образ компании-продавца в сознании такого потребителя приобретет негативный оттенок.

Наилучший выбор форм рекламы для данной группы – реклама в прессе, которую читает именно эта группа потребителей с соответствующим достатком, а также печатная реклама в форме листовок и флаеров и щитовая реклама. Менее действенной рекламой для такой группы потребителей будет реклама в интернете и социальных сетях.

2. Достаточно простые системы скидок с дифференциацией по группам потребителей/клиентов. Излишнее усложнение дисконтных систем, систем накопления баллов недопустимо для некоторых заведений, продвигающих свою миссию, в частности безмятежного отдыха в кафе. Многоуровневые системы скидок – в большей степени пережиток 2000–2008 гг., когда сравнительно небольшое количество компаний с высокой лояльностью потребителей использовала такие системы для удержания постоянных и привлечения новых клиентов.

Верным решением также будет разработка нескольких доступных систем скидок для разных групп потребителей с простейшим алгоритмом накопления бонусов или без него. К примеру, молодежь становится владельцами карт «Активный», дающих скидки на одни группы товаров, а клиентам более старшего возраста предлагается пакет «Все включено», дающий право на фиксированную скидку на все группы товаров, а также большую скидку на конкретный раздел меню на выбор клиента.

Безусловно, любые рекомендации подобного характера должны находиться в прямой зависимости от целого комплекса факторов и могут не подходить, к примеру, под формат того или иного ресторана или кафе. Наиболее сильное влияние на принятие маркетинговых решений окажет характеристика групп клиентов, социальные портреты посетителей.

Перейдем непосредственно к следующему этапу – сегментации накопленных данных.

В данном случае необходимо представить совокупность из  $N$  клиентов, где  $N$  может достигать значений в десятки тысяч, приемлемым для дальнейшего рассмотрения количеством подгрупп. Эффективным методом решения данной проблемы является кластерный анализ.

Название кластерный анализ происходит от английского слова cluster (гроздь, скопление). Задача этого метода заключается в том, чтобы на основании данных, которые содержатся в некотором множестве (в данном случае в таблице наблюдения за посетителями кафе), разбить множество объектов (посетителей) на целое число кластеров (подгрупп), таким образом, чтобы каждый объект принадлежал только одному кластеру и чтобы объекты, принадлежащие одному и тому же кластеру, были схожими, а объекты из разных кластеров – отличались [3].

Кластерный анализ позволяет рассматривать большие объемы информации, сокращая их до компактного, наглядного состояния. Основные его преимущества – это отсутствие каких-либо ограничений, накладываемых на природу рассматриваемых данных, значения могут быть как количественными, так и порядковыми или дихотомическими (признаковыми), а также возможность проводить разбиение объектов по целому набору признаков, а не только одному параметру.

Для определения схожести объектов необходимо использовать некоторую единицу измерения (меру сходства). Наиболее наглядным и широко применяемым методом является использование в качестве этой меры расстояние между объектами в пространстве измеренных параметров. Объекты с меньшим расстоянием между ними являются более похожими, чем объекты с большим расстоянием. При этом существует несколько способов вычисления этого расстояния. Самым широко распространенным является евклидово расстояние, определяемое как:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2},$$

где  $d_{ij}$  – расстояние между объектами  $i$  и  $j$ , а  $x_{ik}$  – значение  $k$  переменной для  $i$ -го объекта. Существуют и другие виды расстояния, например манхэттенское расстояние (или city-block):

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}|,$$

или расстояние Чебышева (супремум-норма), максимальная абсолютная разность в значениях для любой переменной:

$$d_{ij} = \sup_k \{|x_{ki} - x_{kj}|\}, k = 1, 2, \dots, p.$$

При использовании расстояния в качестве меры сходства часто требуется нормировать переменные к нулевому среднему и единичной дисперсии, поскольку переменные с большими абсолютными значениями могут подавить влияние переменных с малыми значениями. Дальнейшие сложности возникают в случаях, когда переменные имеют бинарный характер. В такой ситуации требуется применение коэффициентов ассоциативности в качестве меры сходства. Когда данные имеют смешанный характер (количественные, порядковые и дихотомические) для формирования матрицы сходства требуется применение коэффициента Гауэра:

$$s_{ij} = \frac{\sum_{k=1}^p S_{ijk}}{\sum_{k=1}^p W_{ijk}},$$

где  $W_{ijk}$  – весовая переменная, принимающая значение единицы, если сравнение объектов по признаку  $k$  следует учитывать, и ноль – в противном случае;  $S_{ijk}$  – «вклад» в сходство объектов, зависящий от того, учитывается ли признак  $k$  при сравнении объектов  $i$  и  $j$  [4].

Для двоичных данных  $S_{ij}$  вычисляется в соответствии со следующей системой подсчета (табл. 2).

Таблица 2. Расчет коэффициента Гауэра

Объект $i$	1	1	0	0
Объект $j$	1	0	1	0
Вклад $S_{ijk}$	1	0	0	0
Вес $W_{ijk}$	1	1	1	0

Для порядковых данных  $S_{ijk}$  равно 1, если сравниваемые значения равны, и 0 если нет.

Для количественных данных имеет место уравнение:

$$S_{ijk} = 1 - \frac{|x_{ik} - x_{jk}|}{R_k},$$

где  $x_{ik}$  – значение  $k$  переменной для объекта  $i$ ;  $R_k$  – размах значений этой переменной.

Следующий шаг в кластерном анализе после выбора меры сходства – это определение метода кластеризации. Разработанные кластерные методы на данный момент образуют семь семейств [5]:

- иерархические агломеративные методы;
- иерархические дивизимные методы;
- итеративные методы группировки;
- методы поиска модальных значений плотности;
- факторные методы;
- методы сгущений;
- методы, использующие теорию графов.

Наиболее часто в социальных задачах используются иерархические агломеративные методы. Эти методы просматривают матрицу сходства и последовательно объединяют схожие объекты. В итоге получаем один кластер, объединяющий все рассматриваемые объекты. Важное преимущество иерархических методов заключается в возможности представления процесса объединения в виде древовидной диаграммы (дендограммы) и последующего выбора степени сходства для получения желаемого количества кластеров. Среди них можно выделить метод одиночных связей, метод средних связей, метод полных связей, метод Уорда и др. Использование различных методов может привести к различным результатам разбиения.

Воспользуемся методом Уорда, особенностью которого является стремление к разбиению на кластеры схожего размера. Метод оптимизирует минимальную дисперсию внутри кластеров. Целевой функцией является сумма квадратов отклонений (СКО):

$$СКО = x_j^2 - \frac{1}{n(\sum x_i^2)},$$

где  $x_j$  – значение признака  $j$ -го объекта.

Rescaled Distance Cluster Combine

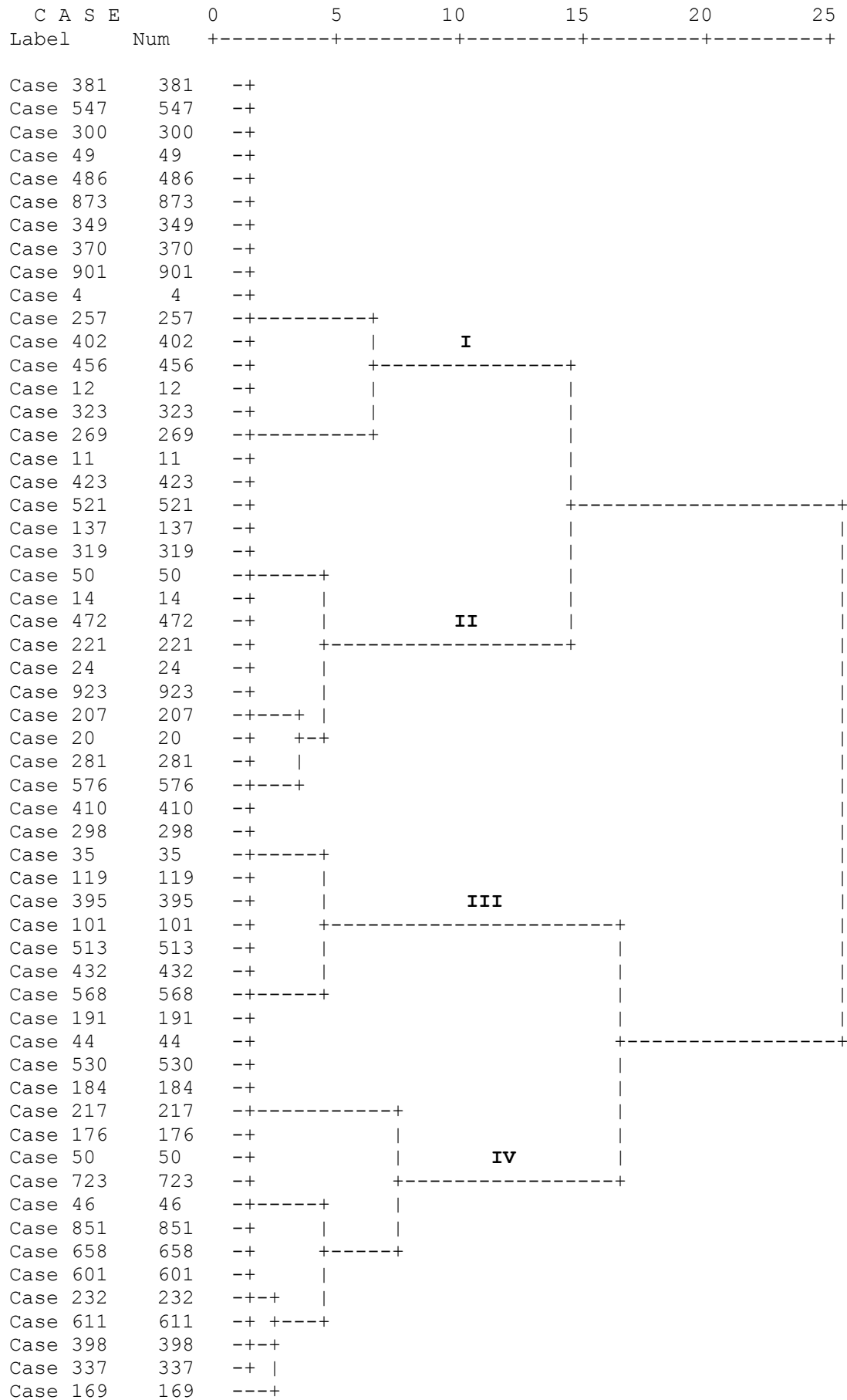


Рис.

Обработка данных заключается в объединении кластеров таким образом, чтобы это вызывало минимальное увеличение их внутригрупповой СКО. На первом шаге все кластеры состоят из одного объекта и их СКО=0. Далее число кластеров последовательно уменьшается, и в итоге все объекты объединяются в один кластер.

Для примера приведено разбиение массива данных о 957-ми посетителях кафе. Наглядно оно представлено следующей дендограммой (показан конечный этап кластеризации) (рис.).

В результате по приращению целевой функции можно выделить четыре кластера:

1. В первый, наиболее многочисленный кластер попала молодежь (в основном учащиеся), некурящие и не употребляющие алкоголь. Средний чек по этой группе равен 333 руб.

2. Второй кластер состоит из курящих людей, в основном женщин, которые пришли без детей. Средний чек 424 руб.

3. Третий, самый маленький кластер – пришедшие с детьми взрослые люди, большинство из них женщины, абсолютно все не употребляют алкоголь и не курят. Средний чек 690 руб.

4. Последний, четвертый кластер представлен взрослыми людьми, употребляющими алкоголь. Большинство из них мужчины, в основном не курящие и без детей. Средний чек 663 руб.

На основании результатов кластерного анализа можно сформулировать следующие рекомендации:

1. Так как большая часть клиентов кафе – это некурящие и не употребляющие алкоголь молодые люди, то при прочих равных условиях, руководству кафе «Суббота» следует инвестировать существенную часть средств, выделенных на рекламу, в продвижение официального сайта компании (SEO-оптимизация, реклама Яндекс-директ, Google Adwords, контекстная реклама, покупка платных ссылок), официальных групп в популярной социальной сети «В Контакте» (SMM-продвижение), а также (после проведения соответствующего функционально-стоимостного анализа) инвестировать средства в разработку мобильной версии сайта или приложения для смартфонов и планшетных ПК, для удобства взаимодействия клиентов с сайтом кафе с мобильных устройств [6].

Однако необходимо грамотно оценить полезный эффект от вышеперечисленных нововведений, так как у данной группы потребителей довольно низкий средний чек. Возможно, внедрение специализированной системы скидок для данной группы клиентов может повлиять на изменение количества посетителей данной группы и на их средний чек.

2. Наиболее высокий средний чек у наименее многочисленной группы – женщины с детьми. При необходимости стимулирования именно этой группы клиентов следует формировать выгодные предложения с иным размером порций для детей – детское меню. В глазах потребителей непривлекательный товар в слишком большой для ребенка порции может преобразоваться в единицу детского меню, имеющую важные для сопровождающего ребенка взрослого отличия от оригинала взрослого меню, и делающего его более привлекательным для заказа. Соответственно, один товар продвигается за счет выгодного отличия от второго.

Целесообразно также использование недорогих, но запоминающихся детям шагов: подарков, использование карандашей для рисунков, телевизора с мультфильмами и т.д.

Но необходимо помнить, что данная группа клиентов наименее многочисленна, и нерационально инвестировать большие средства в развитие этой группы, если нет соответствующего ожидаемого полезного эффекта (экономического, социального).

3. Кластер № 4 представляет клиентов кафе, которые употребляют алкоголь. Необходимо проведение дополнительного исследования для выявления популярных блюд, которые приобретаются совместно с алкоголем или «к алкоголю», а также блюд, от которых посетители отказываются по тем или иным причинам (высокая цена, малая порция и др.). На основе полученных результатов сформулировать рекомендации по продвижению

непопулярных блюд без ущерба или с минимальным в продажах популярных блюд. Не вызывает сомнения также целесообразность специализированного обучения персонала методам эффективной продажи в отношении пьющих клиентов.

4. Большинство клиентов являются некурящими людьми, а доля кластера № 2 является не самой существенной. Необходимо проанализировать процентную заполняемость мест для курящих с целью оптимизации соотношения количества посадочных мест для курящих и некурящих посетителей.

Приведенное исследование построено на базе сети кафе, но его результаты могут быть заметно расширены. Представляется возможным использование кластерных процедур в задачах типизации объектов в интересах ГПС МЧС России, в частности при выборе типовых объектов, для проведения тренировок и учений личного состава.

### **Литература**

1. Траут Дж., Райс Эл. Маркетинговые войны / пер. с англ. СПб.: Питер, 2008.
2. Бугаев Л. Мобильный маркетинг: Как зарядить свой бизнес в мобильном мире. М.: Альпина Паблишер, 2012.
3. Дюран Б., Оделл П. Кластерный анализ. М.: Статистика, 1977.
4. Гайдышев И. Анализ и обработка данных: спец. справ. СПб.: Питер, 2001.
5. Факторный, дискриминантный и кластерный анализ / Дж.О. Ким [и др.]. М.: Финансы и статистика, 1989.
6. Ашманов И., Иванов А. Оптимизация и продвижение сайтов в поисковых системах. 3-е изд. СПб.: Питер, 2013.