



Дьячкова О. В.

МОДЕЛИРОВАНИЕ ФОРМИРОВАНИЯ ДОВЕРИЯ ПОТРЕБИТЕЛЕЙ НА ОСНОВАНИИ КЛАСТЕРНОГО АНАЛИЗА ДАННЫХ

Рассматривается механизм сегментации потребителей на основе применения метода кластеризации k -means. Приведены обоснование выбора метода, разбиение исходных данных на обучающее и тестовое множества, выбор входных и выходных полей для итерационного процесса кластеризации. Для каждого из полученных кластеров проведен анализ их характеристик и дана содержательная их интерпретация. На основании полученных данных выработаны рекомендации PR-стратегии по формированию доверия для каждой целевой группы потребителей.

Ключевые слова: кластерный анализ, сегментация, профиль кластера, мощность кластера, доверие потребителей, PR-стратегия.

1. Введение

Ключевым элементом долгосрочного успеха предприятия является высокого уровня лояльность потребителей [1]. Необходимое условие его достижения — ориентация на клиентов. Для этого следует определить потребности и пожелания потребителей, проанализировать уровень их удовлетворенности от продукции предприятия, их ассоциации, установки и т. д. [2]. На основании данного анализа возможно построить стратегию формирования позитивного отношения и лояльности потребителей.

Для создания положительного имиджа и репутации фармакологической компании поставлена задача формирования доверия целевой аудитории к выпускаемым лекарственным средствам (ЛС). С этой целью следует выделить группы потребителей, на которые направлять максимальные усилия для целенаправленной PR-активности. Этим обосновывается актуальность проведенного исследования.

2. Анализ литературных данных и постановка проблемы

Вопрос, какие именно факторы способствуют формированию восприятия бренда, по-прежнему не имеет четкого ответа. Этот вопрос возникает при использовании как традиционных методов маркетинговых исследований [3], так и новых нейромаркетинговых и иных технологий [4]. Тот же вопрос остается открытым и в случае применения методов нечеткой классификации [5].

Однако проблема не только в выборе переменных для сегментации и определении дифференцирующих критериев. На основании результатов исследования особенностей поведения потребителей ЛС необходимо сформировать механизм удовлетворения их потребностей с использованием инструментов маркетинга [6]. В частности, определение характера пользовательских предпочтений позволит выработать стратегию управления конкурентными преимуществами ЛС на фармацев-

тическом рынке и стратегию управления поведением потребителей [7].

Таким образом, для эффективного решения указанных задач следует провести сегментацию объектов (потребителей) по различным признакам (критериям). Для каждой выделенной группы схожих объектов необходимо затем изучить ее особенности и разработать для нее отдельную стратегию взаимодействия. Поскольку классы объектов (потребителей) заранее неизвестны, используем кластеризацию — разбиение множества объектов на группы, в которых объекты максимально похожи друг на друга и максимально отличаются от объектов из других кластеров [8].

3. Объект, цель и задачи исследования

Объектом исследования является поведение потребителей ЛС определенной торговой марки (компании).

Цель исследования — выработка стратегии формирования позитивного имиджа компании.

Для достижения поставленной цели необходимо выполнить следующие задачи.

Следует сформировать выборку объектов (потребителей) для проведения кластерного анализа.

Необходимо определить множество критериев (признаков), по которым будет проведена кластеризация.

На следующем этапе необходимо провести многомерный статистический анализ исходных данных, а именно — разбить множество исследуемых потребителей на группы однородности. Для каждой группы следует дать ее характеристику и на ее основе выработать план PR-продвижения.

4. Методы исследования

С целью решения поставленных задач было проведено анкетирование среди клиентов компании и потенциальных потребителей ее продукции. В нем приняли участие 300 человек, среди них 246 (82 %) знакомы с препаратами фирмы. В ходе анкетирования было

проведено оценивание уровня доверия потребителей к ЛЗ по 4-ранговой шкале (рис. 1).

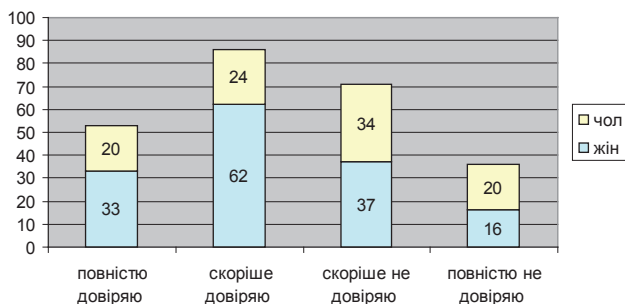


Рис. 1. Уровень доверия потребителей (исходные данные)

В перечень факторов (рис. 2), потенциально влияющих на уровень доверия [9], были включены эффективность, безопасность, ценовая доступность, соответствие цена/качество, оформление и др.

Для кластеризации имеющихся данных был использован пакет Deductor Studio [10]. Эта аналитическая платформа позволяет проводить обработку и анализ структурированных данных с применением Data Mining. Среди обработчиков данного пакета имеется инструмент Кластеризация.

Для проведения кластеризации был выбран алгоритм k -ближайших средних (k -means). В его основе лежит принцип оптимального разбиения множества данных на k кластеров, при котором целевая функция алгоритма достигает экстремума [11]. Выбор метода обусловлен его умеренной вычислительной сложностью с ростом числа записей исходных данных. Также результаты его работы не зависят от порядка записей в исходных данных [12].

Алгоритм состоит из двух этапов.

Этап 1. Первоначальное распределение объектов по кластерам. задается число k , и случайным образом определяется k элементов данных (потребителей). На первом шаге эти точки считаются «центрами» кластеров. Каждому кластеру соответствует один центр. Остальные объекты (потребители) приписываются к ближайшему для каждого из них кластеру.

Этап 2. Итерационный процесс. Вычисляются новые центры кластеров как средние значения каждого признака всех объектов кластера. Объекты перерас-

пределяются по новым кластерам. Процесс вычисления центров и перераспределения объектов продолжается до тех пор, пока не стабилизируются центры кластеров, т. е. после очередного шага все объекты будут принадлежать тому же кластеру, которому они принадлежали до текущей итерации.

Обозначим через x, y, \dots, z — признаки (факторы доверия потребителей), n — количество записей (потребителей) в кластере. Тогда x_i, y_i, \dots, z_i — признаки i -й записи (потребителя). Для расчета нового центра кластера на 2-м этапе алгоритма используется формула вида:

$$(x, y, \dots, z) = \frac{1}{n} \sum_i x_i + \frac{1}{n} \sum_i y_i + \dots + \frac{1}{n} \sum_i z_i. \quad (1)$$

Чтобы определить, к какому из кластеров следует отнести каждый объект, на каждой итерации вычисляются расстояния между объектами и центрами кластеров [13]. Для расстояния между объектами S и P , имеющими признаки (x_s, y_s, \dots, z_s) и (x_p, y_p, \dots, z_p) соответственно, используется евклидово расстояние:

$$d(S, P) = \sqrt{(x_s - x_p)^2 + (y_s - y_p)^2 + \dots + (z_s - z_p)^2}. \quad (2)$$

В качестве критерия сходимости определяют [14] сумму квадратов ошибок r между центром Q кластера и всеми вошедшими в него записями:

$$r = \sum_i ((x_Q - x_i)^2 + (y_Q - y_i)^2 + \dots + (z_Q - z_i)^2) \rightarrow \min. \quad (3)$$

Таким образом, остановка алгоритма произойдет тогда, когда ошибка r достигнет достаточно малого значения.

Для проведения кластеризации все исходное множество данных было разбито на 2 части — обучающее и тестовое (способ разделения данных — случайным образом). Обучающее множество включает записи (строки с данными о потребителях), которые используются для проведения алгоритма k -means, т. е. итерационного процесса. Размер обучающего множества — 95 % данных (234 строки). Тестовое множество служит для проверки найденных результатов. Размер тестового множества — 5 % (12 строк) данных.

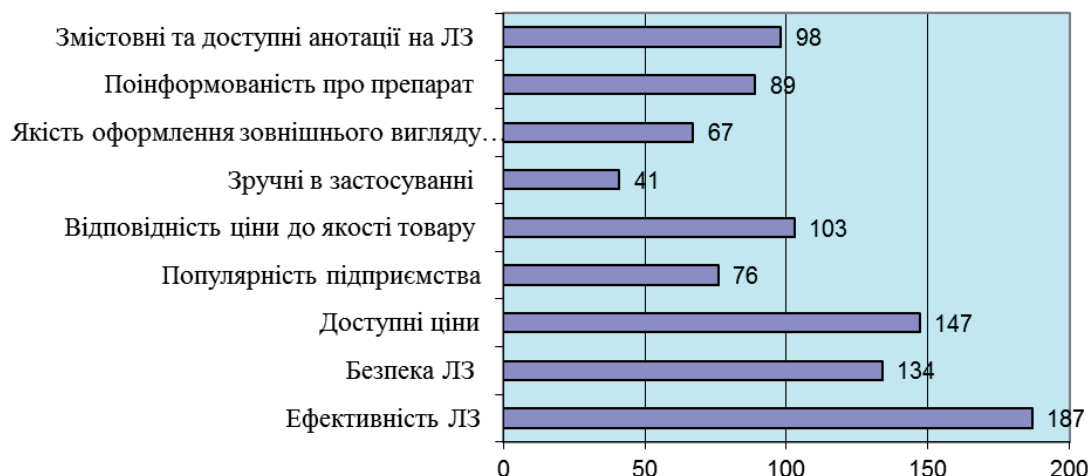


Рис. 2. Факторы доверия (исходные данные)

Оба множества включают записи (примеры), содержащие входные и желаемые целевые значения. В качестве входных полей были указаны поля: *Эффективность ЛС; Безопасность ЛС; Доступные цены; Пол потребителя; Популярность предприятия; Соответствие цена/качество; Удобство применения; Качество оформления; Информированность о препарате; Содержательность и доступность аннотации.* Поля с данными о степени доверия потребителей были указаны как выходные поля.

5. Анализ результатов кластеризации

В результате проведения в Deductor Studio кластерного анализа исходных данных были выделены 4 кластера потребителей. Каждый из них обладает определенным набором характеристик. Это позволяет применить различные подходы к формированию доверия потребителей для различных их групп.

В соответствии с полученной группировкой данные были упорядочены по значимости — степени зависимости результата от полей данных. На рис. 3 значимость данных отображена в числовом и графическом виде

с помощью кросс-таблицы *Профили кластеров* с двумя измерениями *Кластеры* и *Поля*.

Наиболее значимыми факторами для аудитории в целом выявились информированность о ЛС, их безопасность и популярность предприятия (100 % уровень значимости). Также высока значимость доступности цен (77 %).

Для успешного анализа данных [15] следует дать содержательную интерпретацию каждому из полученных кластеров потребителей. Для этого была оценена их мощность, исследован качественный состав, статистические характеристики, распределение значений каждого параметра в кластере.

1. Кластер № 1 (рис. 4) включает 56 потребителей в основном мужского пола (22,8 % от общего числа опрошенных, что составляет 57,2 % всей мужской аудитории). На лояльность этих клиентов к ЛС почти не влияют внешние факторы: информированность о препаратах, популярность предприятия-изготовителя (соответственно 98,9 % и 99,3 %), качество внешнего оформления (73,9 %). Для этой группы потребителей важны содержательные показатели — доступность цен (значимость 86,9 %) и эффективность ЛС (73,6 %).

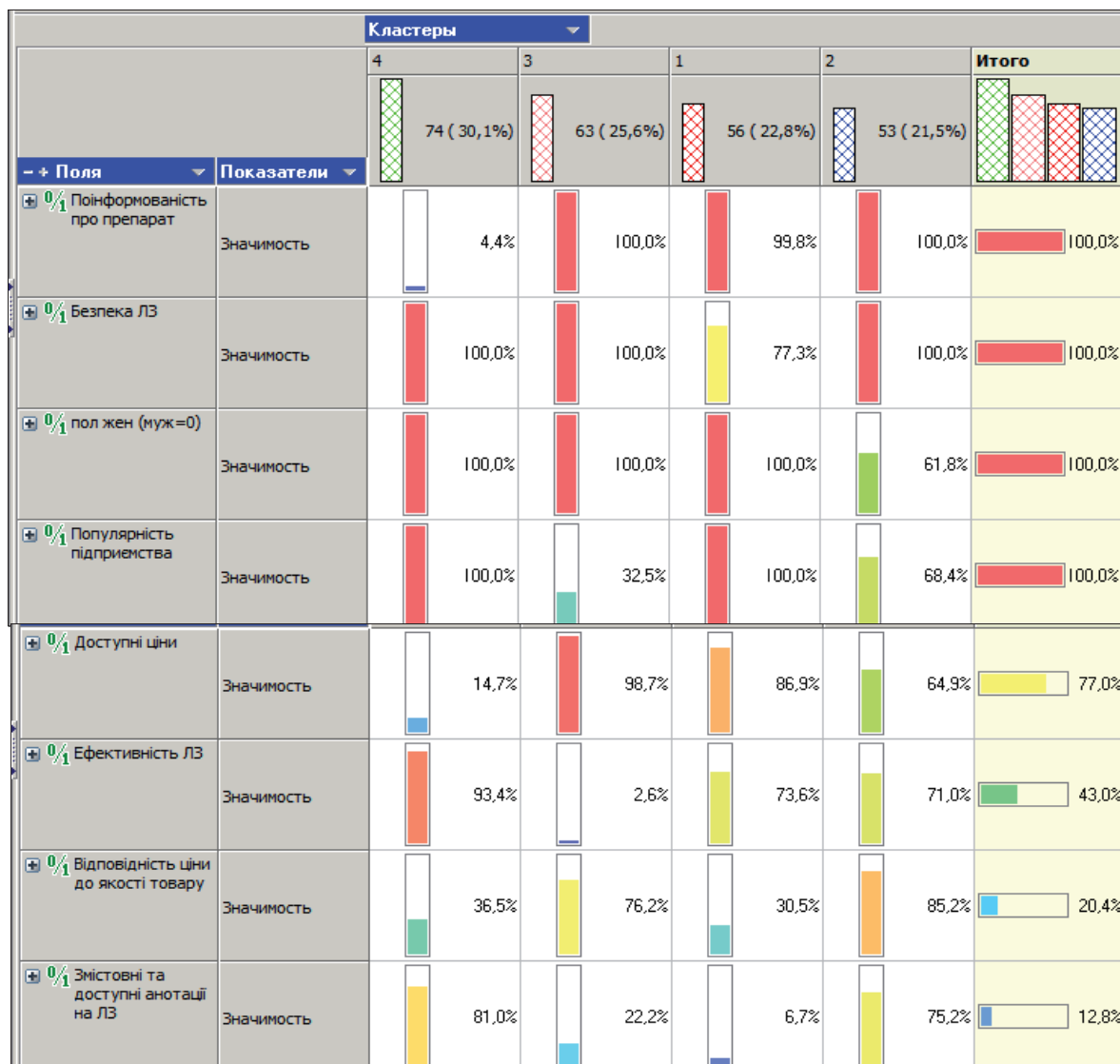


Рис. 3. Результат исследования профилей кластеров

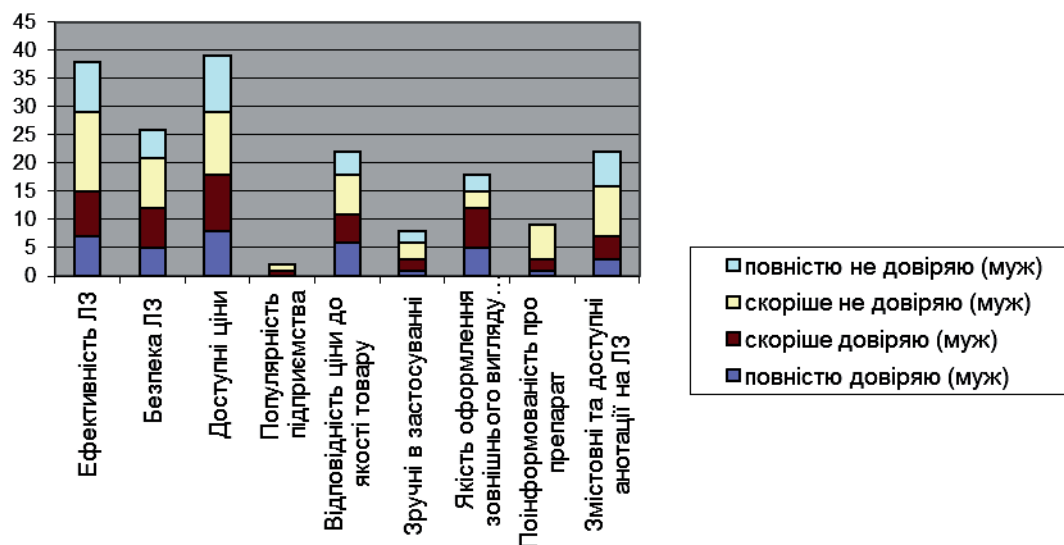


Рис. 4. Значения полей данных Кластера 1

В связи с этим PR-деятельность, направленная на мужскую аудиторию, должна включать логическое обоснование действенности применения препаратов, подчеркивать мягкую ценовую политику. Возможно проведение ценовых акций, скидок и т. п. Не следует для этой целевой группы уделять внимание внешним факторам, поскольку они не влияют на принятие решений.

2. В Кластер № 2 попали 53 потребителя (21 % опрошенных), среди которых женщины составляют 2/3 (66 %), мужчины – 1/3. Для их доверия характерны полная зависимость от информированности о ЛС (значимость 100 %), безопасности и эффективности препаратов (соответственно значимость 100 % и 71 %). Такие факторы, как доступность цен или соответствие цены качеству, имеют меньшую значимость (64,9 %). А 67,9 % потребителей группы даже отметили слабую связь между ценой и качеством препаратов.

Таким образом, для формирования позитивного отношения к ЛС среди этих потребителей следует уделять внимание широкому рекламированию препаратов предприятия, подчеркивая при этом их высокоэффективность и безопасность, что обуславливает соответствующий уровень цен.

3. 63 потребителя Кластера № 3 (25 % опрошенных) не придают особого значения популярности предприятия, оформлению препарата, удобству его использования. Среди тех из них, кто полностью или частично доверяют ЛС, значимость этих факторов составляет 32,5 %, 45,9 %, 13,4 % соответственно. Для них наибольшее значение имеет безопасность препаратов. Однако среди полностью не доверяющих продукции никто не назвал информированность о ней или удобство в использовании, и только 28,3 % отметили качество ее оформления. Отношение к цене у данных потребителей варьирует. В связи с этим сложно сделать вывод о ценовой политике для данной категории. Следует отметить в основном женский состав этой группы (84 %).

Следовательно, для формирования приверженности этих потребителей можно предложить рекламную кампанию, нацеленную на женскую аудиторию, широко информирующую о ЛС и акцентирующую внимание на безопасности.

4. Наконец, 4-й кластер потребителей включает 74 клиента (34,1 %). Для них несущественен фактор безопасности препаратов, но чрезвычайно важна их эффективность (значимость 93,4 %), а также содержательность аннотаций к ним (81,0 %). Информированность о препарате, так же как и цены, не играют роли для данной категории (значимость всего 4,4 % и 14,7 % соответственно).

Таким образом, для данной группы потребителей нет основания для широкой пропагандистской кампании или льготных ценовых акций. Имеет смысл уделить особое внимание составлению описаний и аннотаций, включая упаковку препаратов, подчеркивая в них вопросы эффективности применения.

6. Выводы

В результате проведенных исследований были получены следующие результаты:

Сформирована выборка потребителей ЛС и проведено оценивание уровня их доверия к ЛС по 4-ранговой шкале. Выбраны факторы, потенциально формирующие доверие к ЛС.

Проведена кластеризация данных по методу *k-means*; при этом в качестве расстояния между объектами была выбрана евклидова метрика. В результате получены кластеры – однородные группы потребителей, для которых удалось выявить скрытое сходство их потребительского поведения.

В результате проведенной сегментации была изучена типология потребителей и установлено, что на рынке рассматриваемых ЛС можно выделить 4 группы потребителей. Для каждой из них выявлены внутренние закономерности.

На основании полученных данных выработаны прицельные рекомендации по формированию доверия общественности к продукции предприятия для каждой целевой группы.

Полученные результаты позволяют выстраивать грамотную, целенаправленную PR-активность для максимально эффективного создания положительного имиджа и репутации компании.

Литература

1. Гембл, П. Маркетинг взаимоотношений с потребителями [Текст]: пер. с англ. / П. Гембл, М. Стоун, Н. Вудкок. — М.: ФАИР-ПРЕСС, 2002. — 512 с.
2. Наумова, О. Е. Формування лояльності споживачів як стратегічний напрям менеджменту підприємства [Текст] / О. Е. Наумова // Вісник Хмельницького національного університету. — 2011. — Т. 4, № 6. — С. 171–174.
3. Яременко, С. С. Маркетингові дослідження вподобань споживачів біжутерії на українському ринку [Текст] / С. С. Яременко // Європейський вектор економічного розвитку. — 2014. — № 1(16). — С. 210–216.
4. Пашук, Л. Перспективні напрями використання нейромаркетингу в управлінні поведінкою споживачів [Текст] / Л. Пашук // Маркетинг в Україні. — 2014. — № 3. — С. 16–26.
5. Hiziroglu, A. Customer portfolio analysis: Crisp classification versus fuzzy classification — Based on the supermarket industry [Text] / A. Hiziroglu, J. Patwa, V. Talwar // Journal of Targeting, Measurement and Analysis for Marketing. — 2012. — Vol. 20, № 2. — P. 67–83. doi:10.1057/jt.2012.5
6. Слободянюк, М. М. Структура споживання та споживацькі переваги антигістамінних лікарських препаратів в Україні [Текст]: матеріали наук.-практ. конф. / М. М. Слободянюк, О. С. Самборський // Ефективність використання маркетингу та логістики фармацевтичними організаціями. — Х.: НФаУ, 2008. — С. 114–119.
7. Пестун, І. В. Маркетингове управління поведінкою споживачів, які займаються самолікуванням [Текст] / І. В. Пестун // Запорозький медичний журнал. — 2010. — Т. 12, № 1. — С. 77–81.
8. Horn, B. Comparison of Segmentation Approaches [Electronic resource] / B. Horn, W. Huang // Decision Analyst. — Available at: \www/URL: http://www.decisionanalyst.com/publ_art/CompareSegmentation.dai
9. Хилл, Н. Измерение удовлетворенности потребителей по стандарту ISO 9000:2000 [Текст]: пер. с англ. / Н. Хилл, Б. Сельф, Г. Роше. — М.: Технологии, 2004. — 192 с.
10. Deductor. Руководство аналитика. Версия 5.3 [Текст]. — М.: BaseGroup Labs, 2013. — 213 с.
11. Мандель, И. Д. Кластерный анализ [Текст] / И. Д. Мандель. — М.: Финансы и статистика, 1988. — 176 с.
12. Aggarwal, C. Data Clustering: Algorithms and Applications [Текст] / ed. by C. Aggarwal, C. Reddy. — CRC Press, 2014. — 624 p.
13. Anderberg, M. R. Cluster analysis for applications [Текст] / M. R. Anderberg. — NY: Academic Press, 1973. — 359 p. doi:10.1016/B978-0-12-057650-0.50001-6
14. Bock, H.-H. Origins and extensions of the k-means algorithm in cluster analysis [Electronic resource] / H.-H. Bock // Electronic Journal for History of Probability and Statistics. — December 2008. — Vol. 4, № 2. — P. 1–18. — Available at: \www/URL: http://www.jehps.net/Decembre2008/Bock.pdf
15. Sheth, J. N. Cluster Analysis and its Applications in Marketing Research [Текст] / J. N. Sheth, L. Sherman // Multivariate Methods for Market and Survey Research. — Marketing Classics Press, 2011. — P. 193–208.

МОДЕЛЮВАННЯ ФОРМУВАННЯ ДОВІРИ СПОЖИВАЧІВ НА ПІДСТАВІ КЛАСТЕРНОГО АНАЛІЗУ ДАНИХ

Розглядається механізм сегментації споживачів на основі застосування методу кластеризації k-means. Наведено обґрунтування вибору методу, розбиття вихідних даних на навчальну і тестову множини, вибір вхідних та вихідних полів для ітераційного процесу кластеризації. Для кожного з отриманих кластерів проведено аналіз їх характеристик і дана змістовна їх інтерпретація. На підставі отриманих даних вироблено рекомендації PR-стратегії щодо формування довіри для кожної цільової групи споживачів.

Ключові слова: кластерний аналіз, сегментація, профіль кластера, потужність кластера, довіра споживачів, PR-стратегія.

Дьячкова Ольга Владимировна, кафедра математических методов в экономике, Харьковский национальный университет им. В. Н. Каразина, Украина, e-mail: olga.v.dyachkova@gmail.com.

Дьячкова Ольга Володимирівна, кафедра математичних методів в економіці, Харківський національний університет ім. В. Н. Каразіна, Україна.

Dyachkova Olga, V. N. Karazin Kharkiv National University, Ukraine, e-mail: olga.v.dyachkova@gmail.com