

ТЕХНИЧЕСКИЕ НАУКИ

УДК 004.89

АВТОМАТИЗИРОВАННАЯ СИСТЕМА СБОРА, ОБРАБОТКИ И АНАЛИЗА ДАННЫХ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ РАБОТЫ САЙТОВ

Волкова Алина Александровна, студент, направление подготовки 09.03.01 Информатика и вычислительная техника, Оренбургский государственный университет, Оренбург
e-mail: alinavolk@mail.ru

Сатюков Иван Антонович, студент, направление подготовки 09.03.01 Информатика и вычислительная техника, Оренбургский государственный университет, Оренбург
e-mail: Satyukov77777@mail.ru

Четвергов Алексей Александрович, студент, направление подготовки 09.03.01 Информатика и вычислительная техника, Оренбургский государственный университет, Оренбург
e-mail: chetvergov.w@mail.ru

Научный руководитель: **Семёнов Анатолий Михайлович**, кандидат технических наук, доцент, доцент кафедры программного обеспечения вычислительной техники и автоматизированных систем, Оренбургский государственный университет, Оренбург
e-mail: ng_sem@mail.ru

Аннотация. В настоящей статье авторами представлены результаты исследования и программная реализация модульной архитектуры программного обеспечения автоматизированной системы (АС) сбора, обработки и анализа данных о поведении пользователей, подбор персонализированных виджетов и оценка эффективности сайтов. Каждый программный модуль (ПМ) автоматизированной системы представляет собой завершенное программное обеспечение для решения одной из задач: задачи кластеризации, принятия решений с использованием моделей на основе методов искусственного интеллекта и алгоритмов машинного обучения.

Актуальность исследований обусловлена оценкой конверсии сайтов с целью понимания поведения пользователей сайта и повышением эффективности онлайн-ресурсов. Научная новизна заключается в автоматизации процессов веб-аналитики посредством интеграции отдельных модулей в АС. Дальнейшие исследования могут быть направлены на оптимизацию алгоритмов и расширение функциональности системы.

Ключевые слова: автоматизированная система, пользователи сайта, конверсия, искусственный интеллект, алгоритмы машинного обучения, виджет.

Для цитирования: Волкова А. А., Сатюков И. А., Четвергов А. А. Автоматизированная система сбора, обработка и анализа данных для повышения эффективности работы сайтов // Шаг в науку. – 2025. – № 3. – С. 31–38.

AUTOMATED SYSTEM FOR COLLECTING, PROCESSING AND ANALYZING DATA TO IMPROVE THE EFFICIENCY OF WEBSITES

Volkova Alina Aleksandrovna, student, training program 09.03.01 Computer Science and Computer Engineering, Orenburg State University, Orenburg
e-mail: alinavolk@mail.ru

Satyukov Ivan Antonovich, student, training program 09.03.01 Computer Science and Computer Engineering, Orenburg State University, Orenburg
e-mail: Satyukov77777@mail.ru

Chetvergov Aleksey Aleksandrovich, student, training program 09.03.01 Computer Science and Computer Engineering, Orenburg State University, Orenburg
e-mail: chetvergov.w@mail.ru

Research advisor: **Semyonov Anatoly Mikhailovich**, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Computer Engineering and Automated Systems Software, Orenburg State University, Orenburg
e-mail: ng_sem@mail.ru

Abstract. In this article, the authors present the research results and the software implementation of the modular software architecture of an automated system for collecting, processing and analyzing data on user behavior, selecting personalized widgets and evaluating the effectiveness of websites. Each software module (PM) of an automated system is a complete software for solving one of the tasks: clustering tasks, decision-making using models based on artificial intelligence methods and machine learning algorithms.

The relevance of the research is due to the assessment of website conversions in order to understand the behavior of website users and increase the effectiveness of online resources. The scientific novelty lies in the automation of web analytics processes through the integration of individual modules into the automated control system. Further research may be aimed at optimizing algorithms and expanding the functionality of the system.

Key words: automated system, website users, conversion, artificial intelligence, machine learning algorithms, widget.

Cite as: Volkova, A. A., Satyukov, I. A., Chetvergov, A. A. (2025) [Automated system for collecting, processing and analyzing data to improve the efficiency of websites]. *Shag v nauku* [Step into science]. Vol. 3, pp. 31–38.

Модульная архитектура программного обеспечения АС объединяет три взаимосвязанных ПМ, которые в совокупности позволяют достичь цели исследования: автоматизация процессов сбора, анализа и учета данных о поведении пользователей с целью разделения на кластеры с похожим поведением и характеристиками, что позволяет персонализировать подбор виджетов для каждого пользователя и тем самым повысить конверсию сайта.

Первый модуль. Анализирует модели поведения пользователей и предназначен для сбора и анализа данных о взаимодействии пользователей с сайтом. Фиксирует ключевые метрики, такие как время на странице, глубина просмотра, процент прокрутки, количество сессий и бездействие, позволяя выявлять закономерности поведения и распределять пользователей по кластерам.

Второй модуль. Реализует поддержку принятия решений на основе нечеткого вывода и предназначен для

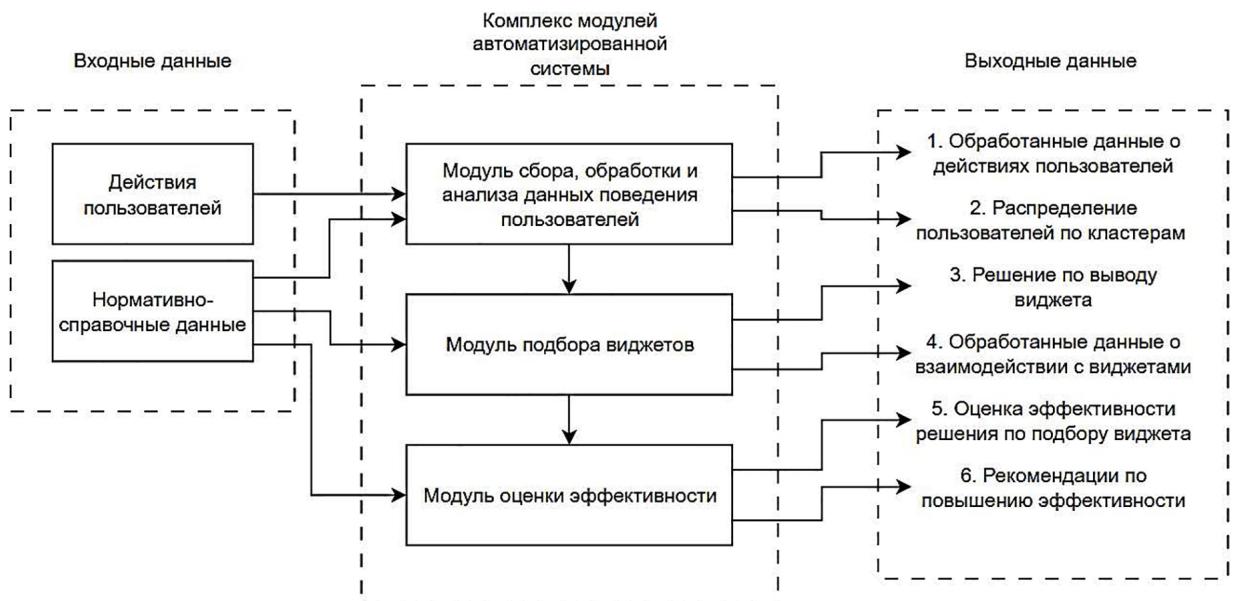


Рисунок 1. Модульная структура автоматизированной системы
Источник: разработано авторами

подбора виджета пользователю на основе данных о его поведении и принадлежности к кластеру, полученному в первом модуле. На основе этой информации разработан прототип базы нечетких правил и реализован алгоритм нечеткого вывода Мамдани, по результатам которого принимается решение о том, какой виджет является наиболее подходящим для пользователя. Данные о взаимодействии пользователей с виджетом (отправка или закрытие формы) передаются в следующий модуль.

Третий модуль. Оценка коэффициента конверсии сайта анализирует эффективность работы веб-ресурса, оценивая, насколько успешно сайт выполняет поставленные бизнес-цели. Позволяет выявлять факторы, влияющие на конверсию, и принимать обосново-

ванные решения по ее повышению. Для оценки коэффициентов конверсии сайтов используется критериальный подход (метод анализа иерархий).

Модульная структура автоматизированной системы представлена на рисунке 1.

Выполнен обзор и анализ аналогов систем веб-аналитики: Google Аналитика, Яндекс Метрика, OptinMonster.

Google Аналитика – это платформа для сбора данных с сайтов и приложений о количестве посетителей, их действиях и создания отчетов, в которых отслеживается путь клиентов и уровень повышения рентабельности инвестиций в маркетинг¹. В качестве примера на рисунке 2 представлено программное средство «Google Analytics».

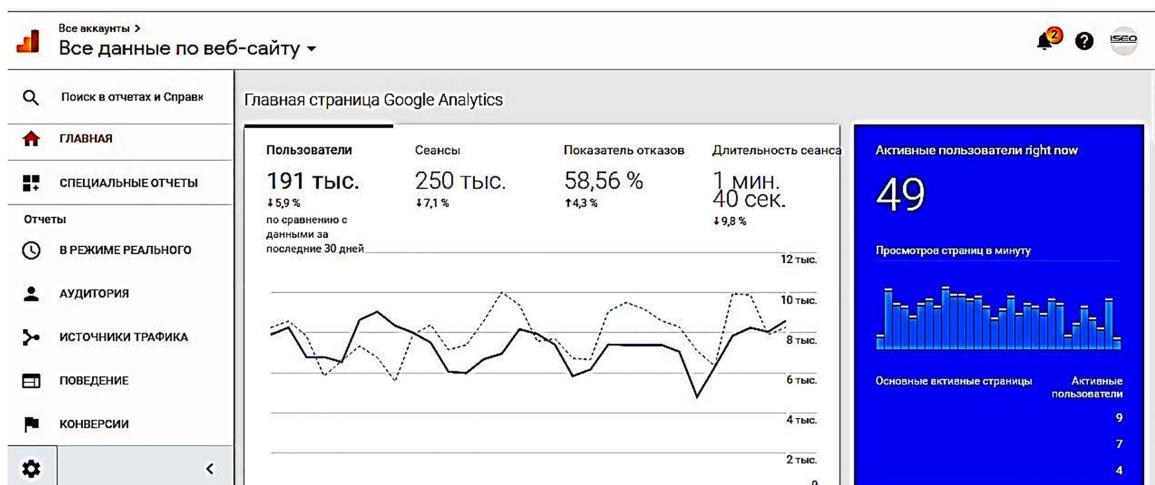


Рисунок 2. ПС «Google Analytics»

Источник: взято из источника «Официальный сайт Google Аналитика»²

Яндекс Метрика – веб-аналитический сервис от Яндекса, предназначенный для отслеживания и анализа поведения пользователей на сайте³.

OptinMonster – инструмент для создания всплывающих окон и слайдеров на сайте. С его помощью можно автоматизировать сбор адресов электронных почт посетителей сайта⁴.

Плюсами рассмотренных программных средств являются: сбор большого объема данных о поведении пользователей; легкая интеграция с маркетинго-

выми инструментами; подробный анализ посетителей; набор инструментов для создания и управления маркетинговыми кампаниями; легкость настройки и использования.

Минусами программных средств являются: платное использование после окончания пробного периода; ограничение некоторых функций в бесплатной версии; ограничения по экспорту данных и интеграции с внешними системами анализа; ограниченный набор шаблонов виджетов; ограниченное количество

¹ Справочный материал программного средства Google Analytic – URL: <https://analytics.google.com/> (дата обращения: 21.04.2025).

² Справочный материал программного средства Google Аналитика – URL: <https://analytics.google.com/> (дата обращения: 24.04.2025).

³ Справочный материал программного средства Яндекс.Метрика – URL: <https://metrika.yandex.ru/> (дата обращения: 21.04.2025).

⁴ Справочный материал программного средства OptinMonster – URL: <https://optinmonster.com/> /docs/welcome-and-overview-of-optinmonster/ (дата обращения: 21.04.2025).

параметров для настройки правил показа.

На основании вышеизложенного, принято решение о разработке приложения, которое позволит автоматизировать процессы сбора данных, кластеризацию пользователей на основе их активности, подбор виджета пользователю на основании его действий и расчёту конверсии.

На этапе анализа поведения пользователей (первый модуль) происходит кластеризация на основе алгоритма *k-means*, позволяющая разделить пользователей на группы (кластеры) для принятия решений

$$c_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_i, \quad (1)$$

где

c_j – новый центройд j -го кластера (вектор той же размерности, что и объекты x_i);
 $|S_j|$ – количество объектов в кластере S_j (мощность множества).

4. Повторение шагов 2 и 3 до тех пор, пока положение центроидов не перестанет существенно изменяться или пока не будет достигнуто заданное количество итераций.

Для оценки качества кластеризации использованы метрики:

1. Silhouette Score (Коэффициент силуэта) – измеряет, насколько хорошо элементы в кластере сгруппиро-

ваны и насколько они отличаются от элементов других кластеров.

2. Сумма квадратов внутрикластерных расстояний (WCSS). Сумма квадратов расстояний между точками и центроидами внутри каждого кластера. Чем меньше эта величина, тем более компактными и раздельными являются кластеры:

$$WCSS = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - c_i\|^2, \quad (2)$$

где

$WCSS$ – сумма квадратов внутрикластерных расстояний;
 k – количество кластеров;
 C_i – множество объектов, принадлежащих i -му кластеру;
 x_j – объект (точка данных) в кластере C_i ;
 c_i – центройд i -го кластера.

После выполнения кластеризации с помощью алгоритма *k-means* для анализа и интерпретации результатов применен метод Principal Component Analysis (PCA). PCA используется для снижения размерности данных, что позволяет визуализировать многомерные данные в двумерном или трехмерном пространстве, сохраняя при этом максимальную информацию о структуре данных [4].

PCA основан на сингулярном разложении или вычислении собственных векторов и собственных значений ковариационной матрицы данных. Главные компоненты – это направления, вдоль которых данные имеют наибольшую изменчивость. Формально, PCA минимизирует ошибку проекции данных на подпространство меньшей размерности:

$$\min \sum_{i=1}^n \|x_i - \hat{x}_i\|^2, \quad (3)$$

где

x_i – исходные данные;
 \hat{x}_i – их проекция на подпространство главных компонент.

Визуализация кластеров, где отображены основные характеристики и их распределение представлены на рисунке 3.

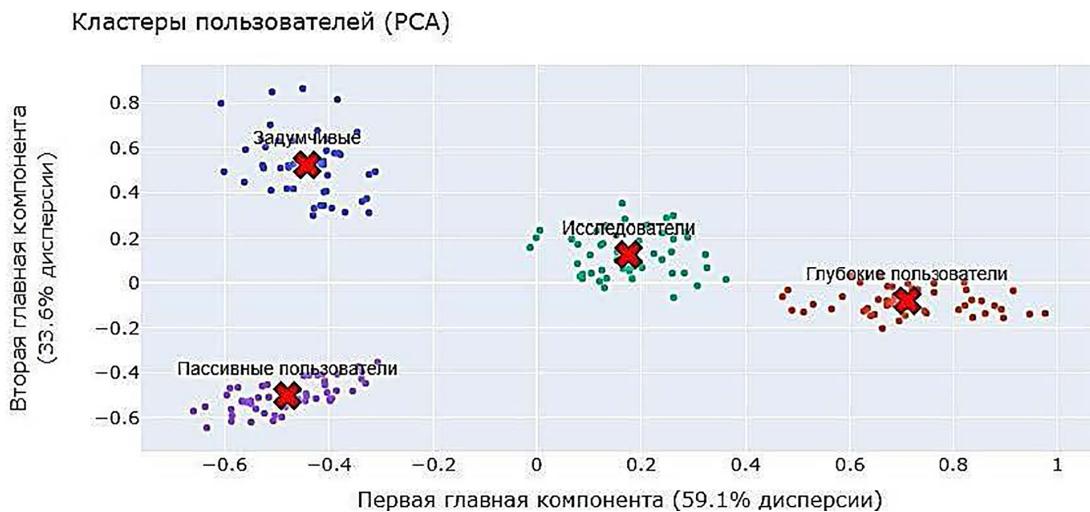


Рисунок 3. Визуализация кластеров

Источник: разработано авторами

Анализ полученных результатов позволяет принять решение о результатах распределения пользователей на кластеры и оценить качество кластеризации по метрикам качества: коэффициент силуэта $K = 0,87$; сумма квадратов внутрикластерных расстояний $WCSS = 10,2$.

На этапе принятия решения (второй модуль) по подбору виджета используется алгоритм нечеткого вывода Мамдани [3]. Этот алгоритм используется для работы с нечеткими данными и позволяет строить логические правила на основе экспертных знаний⁵.

Вид и параметры треугольной функции принадлежности получены авторами на основе метода парных сравнений [7]. В разрабатываемой системе виджет подбирается на основе данных о пользователе.

Значения входных лингвистических переменных:

«Время на сайте» = {«короткое» (1, 1, 150); «среднее» (100, 150, 300); «длительное» (250, 600, 600);}

«Глубина просмотра» = {«маленькая» (1, 1, 5); «средняя» (3, 5, 8); «большая» (6, 10, 10);}

«Процент прокрутки» = {«маленький» (0, 0, 50); «средний» (30, 50, 80); «большой» (70, 100, 100);}

«Кластер» = {«заинтересованных» (1, 1, 2); «лояльных» (1, 2, 3); «социальных» (3, 4, 4);}

Значения выходных лингвистических переменных:

«Виджет» = {«конверсионный» (0, 0, 1); «социальный» (0, 1, 2); «рейтинг» (1, 2, 3); «скидочный» (2, 3, 3)}

Фрагмент прототипа базы нечетких производственных правил представлен в таблице 1.

Таблица 1. Нечеткие производственные правила

	Время на сайте		Глубина просмотра		Процент прокрутки		Кластер		Виджет
ЕСЛИ	короткое	И	большая	И	большой	И	социальных	ТОГДА	рейтинг
ЕСЛИ	среднее	И	большая	И	средний	И	лояльных	ТОГДА	конверсионный
ЕСЛИ	короткое	И	средняя	И	большой	И	заинтересованных	ТОГДА	скидочный
ЕСЛИ	длительное	И	маленькая	И	средний	И	лояльных	ТОГДА	социальный

Источник: разработано авторами

Реализация алгоритма Мамдани и тестирование в среде Fuzzy Logic представлены на рисунке 4.

⁵ Назаров Д. М., Конышева Л. К. Интеллектуальные системы: основы теории нечетких множеств: учеб. пособие для академического бакалавриата. – 3-е изд., испр. и доп. – М.: Издательство Юрайт, 2019. – 186 с.

Ведите параметры пользователя

Время на сайте (секунды): 450
 Глубина просмотра (страницы): 7
 Процент прокрутки: 80
 Кластер (1-4): 2
 Подобрать виджет

Подобранный виджет: rating

Насколько вы согласны с утверждениями ниже?

Сайт работает быстро

1	2	3	4	5
Совсем НЕ соглашаюсь				Полностью соглашаюсь

Мне легко найти нужную информацию

1	2	3	4	5
Совсем НЕ соглашаюсь				Полностью соглашаюсь

Сайт выглядит современным и привлекательным

1	2	3	4	5
Совсем НЕ соглашаюсь				Полностью соглашаюсь

ОТПРАВИТЬ

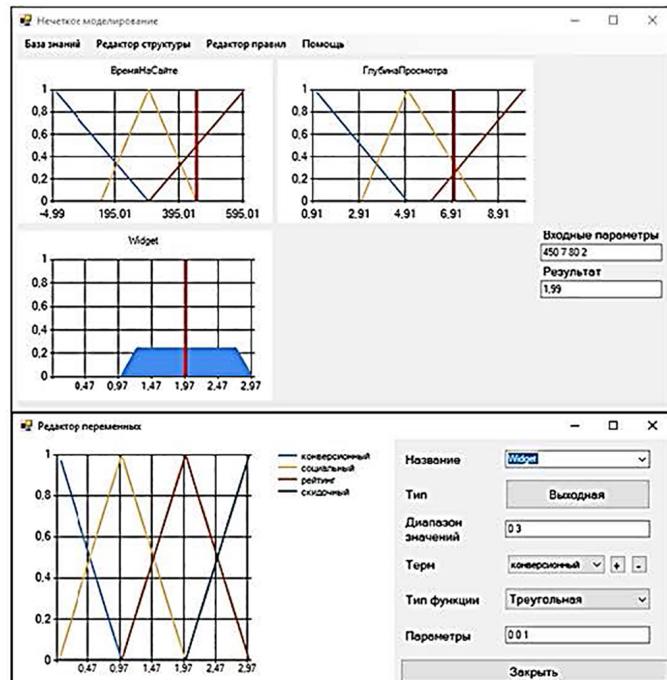


Рисунок 4. Реализация и тестирование алгоритма Мамдани

Источник: разработано авторами в среде Fuzzy Logic

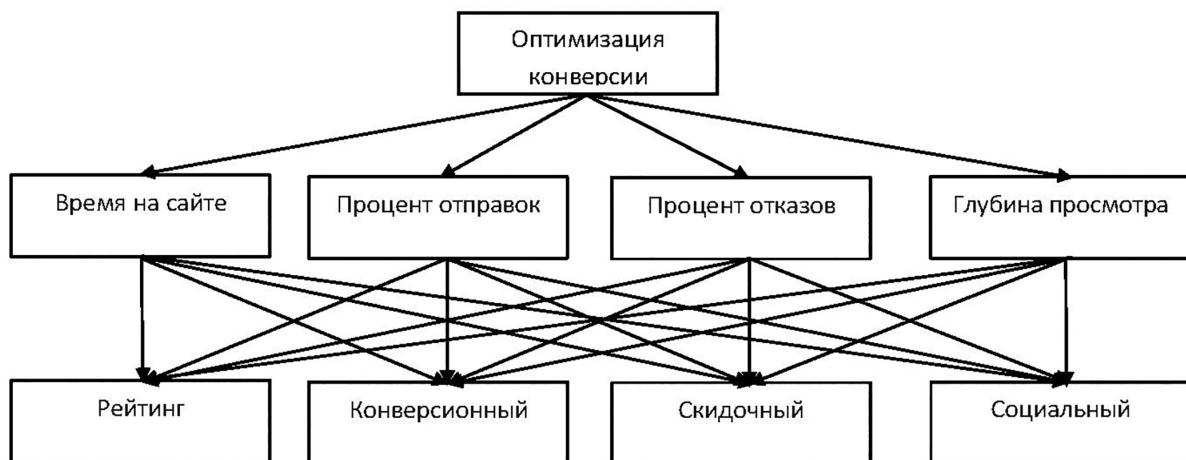


Рисунок 5. Иерархическая модель МАИ

Источник: разработано авторами

Для оценки коэффициентов конверсии сайтов (третий модуль) используется метод анализа иерархий⁶. Этот алгоритм применяется для решения задач многокритериального выбора и позволяет комплексно учитывать взаимное влияние различных параметров

на конечный результат [6]. Иерархическая модель МАИ представлена на рисунке 5.

После определения структуры иерархии выполняется парное сравнение элементов. Пример матрицы парных сравнений представлен в таблице 2.

⁶ Саати Томас Л. Принятие решений: Метод анализа иерархий // Радио и связь. Пер. с англ. Р. Г. Вачнадзе. – М. : Радио и связь, 1993. – 314 с.

Таблица 2. Матрица парных сравнений

Критерий	Время на сайте	Процент отправок	Процент отказов	Глубина просмотра
Время на сайте	1,00	3,00	5,00	2,00
Процент отправок	0,33	1,00	3,00	0,50
Процент отказов	0,20	0,33	1,00	0,25
Глубина просмотра	0,50	2,00	4,00	1,00

Источник: разработано авторами

Для каждого уровня строится своя матрица. После заполнения матрицы второго уровня аналогичным образом формируются матрицы для третьего уровня. Далее вычисляются локальные приоритеты – веса каждого элемента относительно других

в рамках одного уровня.

Построение матрицы парных сравнений 2 уровня и расчёт локальных приоритетов виджетов представлены таблицах 3, 4.

Таблица 3. Матрица парных сравнений

Виджет	Рейтинг	Конверсионный	Скидочный	Социальный
Рейтинг	1.00	0.50	0.33	2.00
Конверсионный	2.00	1.00	0.50	3.00
Скидочный	3.00	2.00	1.00	4.00
Социальный	0.50	0.33	0.25	1.00

Источник: разработано авторами

Таблица 4. Локальные приоритеты

Рейтинг	0.160
Конверсионный	0.278
Скидочный	0.467
Социальный	0.095

Источник: разработано авторами

После определения локальных весов всех уровней выполняется синтез глобальных приоритетов. Пример глобальных приоритетов представлен в таблице 5.

Таблица 5. Локальные приоритеты

Рейтинг	0.138
Конверсионный	0.235
Скидочный	0.389
Социальный	0.238

Источник: разработано авторами

В результате работы модуля автоматизированной системы получен отчёт и рекомендации по улучшению конверсии [5]. Пример выходных данных представлен в таблице 6.

Таблица 6. Выходные данные

Название сайта	Тип сайта	Прошлый коэффициент конверсии	Новый коэффициент конверсии	Изменение конверсии, %	Лучший виджет (% отказов)	Худший виджет (% отказов)
Интернет-магазин электроники	Интернет-магазин	2%	2,5%	+25%	Конверсионный	Рейтинг
Сайт туристического агентства	Информационный сайт	1,5%	1,8%	+20%	Рейтинг	Скидочный
Блог о здоровом образе жизни	Блог	3,0%	3,2%	+6,7%	Скидочный	Социальный

Источник: разработано авторами

Модульный подход к проектированию программного обеспечения позволяет разрабатывать гибкие программные решения, которые могут развиваться по

мере изменения требований. Дальнейшие исследования могут быть направлены на оптимизацию алгоритмов и расширение функциональности системы.

Литература

1. Вахромеева Е. Н., Зензинова Ю. Б. Автоматизация кластеризации компаний по финансовым показателям с использованием k-means алгоритма на больших данных // Дискуссия. – 2024. – № 5(126). – С. 46–50. – <https://doi.org/10.46320/2077-7639-2024-5-126-46-50>.
2. Исаченко Р. В., Катруца А. М. Метрическое обучение и снижение размерности пространства в задачах кластеризации // Машинное обучение и анализ данных. – 2016. – Т. 2, № 1. – С. 17–25.
3. Карамышев А. Н. Основы нечеткой логики / Р. Григорьева, Г. А. Гареева, Р. Р. Басыров – Набережные Челны: Изд-во НЧИ КФУ, 2018. – 11 с.
4. Кисляков А. Н., Тихонюк Н. Е. Выбор метода сегментирования клиентской базы в условиях информационной асимметрии // Вестник Алтайской академии экономики и права. – 2021. – № 2. – С. 46–52. – <https://doi.org/10.17513/vaael.1596>.
5. Плотников В. А. Анализ экономической деятельности интернет-магазина с помощью средств веб-аналитики // Цифровые модели и решения. – 2024. – Т. 3, № 1. – С. 20–30. – <https://doi.org/10.29141/2949-477X-2024-3-1-2>.
6. Саати Т. Л. Принятие решений при зависимостях и обратных связях: Аналитические сети. – М.: ЛКИ, 2008. – 360 с.
7. Чернов В. Г. Нечеткие множества. Основы теории и применения : учеб. пособие. – Владимир: Изд-во ВлГУ, 2018. – 156 с.

Статья поступила в редакцию: 29.04.2025; принята в печать: 02.07.2025.

Авторы прочитали и одобрили окончательный вариант рукописи.