

УДК 336.11

ИСПОЛЬЗОВАНИЕ BIGDATA В МАРКЕТИНГОВОЙ ДЕЯТЕЛЬНОСТИ

Т.А. САХНОВИЧ¹, Б.В. ПОГОДИН², Н.О. КРАСИЛОВ²

¹ к.э.н., доцент кафедры «Инженерная экономика»

² студент учебной группы 10302123

Белорусский национальный технический университет
г. Минск, Республика Беларусь

Аннотация. В современном маркетинге использование больших данных (Big Data) становится ключевым аспектом для достижения конкурентных преимуществ. Анализ больших массивов информации позволяет компаниям глубже понять потребительское поведение, предсказывать тренды и оптимизировать рекламные кампании. С помощью технологий обработки данных маркетологи могут сегментировать аудиторию, персонализировать предложения и улучшать взаимодействие с клиентами. В данной работе рассматриваются основные методы и инструменты анализа больших данных, а также примеры их успешного применения в различных отраслях. Особое внимание уделяется этическим аспектам использования персональных данных и вызовам, связанным с их безопасностью.

Ключевые слова: большие данные, цифровой маркетинг, электронный бизнес.

USING BIG DATA IN MARKETING ACTIVITIES

T.A. SAHNOVICH¹, B.V. POGODIN², N.O. KRASILOV²

¹ Ph.D in Economics, Associate Professor of the Department

«Engineering Economics»

² group student 10302123

Belarusian National Technical University
Minsk, Republic of Belarus

Abstract: In modern marketing, the use of big data is becoming a key aspect for achieving competitive advantages. Analysis of large amounts of

information allows companies to better understand consumer behavior, predict trends and optimize advertising campaigns. With the help of data processing technologies, marketers can segment the audience, personalize offers and improve interactions with customers. This paper discusses the main methods and tools for analyzing big data, as well as examples of their successful application in various industries. Particular attention is paid to the ethical aspects of the use of personal data and the challenges associated with their security.

Keywords: Big data, digital marketing, e-business.

С каждым годом количество информации в сети интернет увеличивается во много раз. Множество источников информации, которые существуют на бумаге переносятся в интернет, бизнес – не исключение. С развитием технологий и цифровизации общества все компании и учреждения стараются автоматизировать свой бизнес и уменьшить время на обработку и анализ информации. Таким образом. Можно сделать вывод, что каждая компания, которая имеет свои инструменты электронного бизнеса так или иначе сталкивалась с BigData. К примеру, данные всех клиентов в базе данных является примером структурированных больших данных, так как данных список скорее всего будет отсортирован либо по алфавитному порядку, либо по дате обращения клиента. Достаточно трудно проанализировать количество всех клиентов, либо количество проданной продукции, когда клиентов десятки тысяч, а на человеческий анализ нет времени, именно для этого применяется технология BigData.

Big Data или большие данные – это структурированные или неструктурированные массивы данных большого объема. Их обрабатывают при помощи специальных автоматизированных инструментов, чтобы использовать для статистики, анализа, прогнозов и принятия решений (рисунок 1).

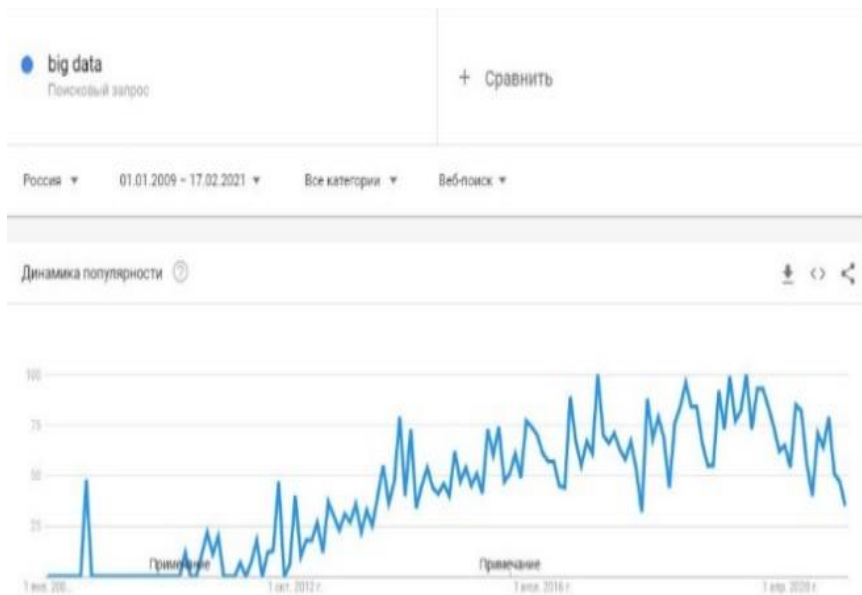


Рисунок 1 – Всплеск интереса к большим данным в Google Trends

Использование Big Data охватывает широкий спектр областей и приложений, которые значительно изменяют подход к анализу информации и принятию решений.

Одним из основных направлений является бизнес-аналитика. Компании используют большие объемы данных для выявления трендов, анализа поведения клиентов и оптимизации своих предложений. Это позволяет им более точно таргетировать рекламу и улучшать качество обслуживания.

Используя большие данные в современных маркетинговых стратегиях, компании и бизнес могут добиться большого влияния в области привлечения клиентов, их удержания, оптимизации расходов на маркетинг, визуализации данных, разработки новых маркетинговых стратегий, прогнозирования продаж, снижение коэффициента оттока клиентов.

Исходя из множества научных исследований, можно понять, что визуальная информация воспринимается человеком во много раз лучше и эффективнее в последующем запоминании по сравнению с другими форматами информации. Во время работы с большими объемами данных, крайне важно визуализировать информацию для более эффективной работы с ней. Примерами визуализации являются схемы, диаграммы, интеллект-карты (рисунок 2).



Рисунок 2 – Данные в виде схем и графиков

Большие данные в маркетинге играют огромную роль, но работа с ними требует конкретных задач, которые директора по маркетингу хотят извлечь из массива информации. Именно качественно разработанная стратегия работы с большими данными и постановка целей, разбитие конечной цели на подзадачи позволяет достичь наилучший результат после анализа.

Работа с Биг Дата в маркетинговых агентствах имеет свою специфику, которую важно учитывать.

Сбор больших данных для маркетинга происходит через разнообразные каналы, включая социальные сети, CRM-системы, мобиль-

ные приложения, веб-аналитику, онлайн-опросы и офлайн-источники. Важно собирать информацию о клиентах, начиная с их поведения на сайте и заканчивая историей покупок.

Процесс сбора больших данных не осуществляется вручную; для этого применяются автоматизированные решения, которые извлекают данные из различных источников и хранят их в базе, где затем начинается работа с ними с использованием искусственного интеллекта. Однако настройка быстрой и корректной передачи данных представляет собой серьезную задачу, для решения которой обычно привлекаются data-инженеры, работающие как на аутсорсе, так и в штате. Они, основываясь на техническом задании от маркетологов, выбирают источники данных и настраивают интеграции, чтобы база данных постоянно обновлялась.

Следует отметить, что в рамках Big Data содержится разнообразная информация, такая как фотографии и видео, сообщения, отзывы, хештеги и другая текстовая информация, данные профилей в социальных сетях, истории действий пользователей и информация о персональных профилях. Выбор конкретных данных зависит от задач, стоящих перед маркетологами. Например, для создания персонализированных рекомендаций могут учитываться последние покупки, оценки товаров, добавления в «Корзину» и средняя стоимость заказов.

Хранение больших данных (Big Data) связано с обработкой огромных объемов информации, высокой скоростью обработки и разнообразием форматов, что делает невозможным их хранение на традиционных устройствах. Для решения этой задачи применяются специализированные инструменты, такие как распределенные файловые системы (например, HDFS в Apache Hadoop), облачные хранилища (Amazon S3, Google Cloud Storage) и гипермасштабируемые базы данных (Snowflake, Cassandra).

BigData имеют классификацию по уровню структурированности данных, насчитывают 3 типа больших данных: структурированные, полуструктурированные, не структурированные.

Структурированные данные – из названия можно понять, что эти данные имеют определенную структуру, как правило в виде таблиц. Структурированные данные пользователи могут делать при помощи специальных приложений, например, MS Excel (рисунки 3).

Employee_ID	Employee_Name	Gender	Department	Salary_In_lacs
2365	Rajesh Kulkarni	Male	Finance	650000
3398	Pratibha Joshi	Female	Admin	650000
7465	Shushil Roy	Male	Admin	500000
7500	Shubhojit Das	Male	Finance	500000
7699	Priya Sane	Female	Finance	550000

Рисунок 3 – Пример структурированных данных в виде таблицы

Полуструктурированные данные – это данные, которые не имеют табличного вида, но при этом имеют некоторые разграничители, которые разделяют данные на блоки или классы. К примеру, html код имеет в своей структуре главные атрибуты, такие как head и body, а все остальные компоненты входят в состав этих блоков (рисунок 4).

```
<rec><name>Prashant Rao</name><sex>Male</sex><age>35</age></rec>
<rec><name>Seema R.</name><sex>Female</sex><age>41</age></rec>
<rec><name>Satish Mane</name><sex>Male</sex><age>29</age></rec>
<rec><name>Subrato Roy</name><sex>Male</sex><age>26</age></rec>
<rec><name>Jeremiah J.</name><sex>Male</sex><age>35</age></rec>
```

Рисунок 4 – Пример полуструктурированных данных

Неструктурированные данные представляют собой информацию, которая не имеет predefined модели или формата, к примеру такими являются отзывы и рейтинги, фото и видео, социальные медиа.

Компания Meta Group предложила основные характеристики больших данных:

Volume – объем данных: от 150 Гб в сутки;

Velocity – скорость накопления и обработки массивов данных. Большие данные обновляются регулярно, поэтому необходимы интеллектуальные технологии для их обработки в режиме онлайн;

Variety – разнообразие типов данных. Данные могут быть структурированными, неструктурированными или структурированными частично. Например, в соцсетях поток данных не структурирован: это могут быть текстовые посты, фото или видео.

Veracity – достоверность как самого набора данных, так и результатов его анализа;

Variability – изменчивость. У потоков данных бывают свои пики и спады под влиянием сезонов или социальных явлений. Чем нестабильнее и изменчивее поток данных, тем сложнее его анализировать;

Value – ценность или значимость. Как и любая информация, большие данные могут быть простыми или сложными для восприятия и анализа. Пример простых данных – это посты в соцсетях, сложных – банковские транзакции

С целью обеспечения масштабируемости и отказоустойчивости данные часто хранятся в распределенной инфраструктуре, где информация делится на блоки и распределяется по различным серверам. Эта архитектура снижает риск потери данных и ускоряет их обработку. Учитывая огромные объемы Big Data, важно применять технологии сжатия (compression) для уменьшения занимаемого пространства и методы дедупликации для исключения дублирующихся данных.

Эффективное взаимодействие с Big Data требует комплексного подхода, объединяющего хранение, обработку, анализ и визуализацию данных, а также надежную систему безопасности. Простая интеграция хранилища данных с инструментами анализа и визуализации является лишь началом пути к успешному использованию больших данных. Рассмотрим подробнее каждый из аспектов.

Выбор хранилища зависит от специфики данных и задач. Для структурированных данных, таких как данные продаж или финансовые отчеты, подходят реляционные базы данных (RDBMS), например, PostgreSQL или MySQL. Однако, для Big Data, часто характеризующейся огромными объемами и разнообразием форматов, реляционные базы данных оказываются недостаточно эффективными. В та-

ких случаях применяются распределенные системы хранения данных, такие как Hadoop Distributed File System (HDFS) или облачные хранилища, предлагаемые Amazon S3, Google Cloud Storage или Azure Blob Storage. HDFS, например, разбивает большие файлы на блоки и распределяет их по множеству узлов кластера, обеспечивая высокую отказоустойчивость и масштабируемость. Облачные хранилища предоставляют гибкость и масштабируемость «по требованию», освобождая от необходимости управления собственной инфраструктурой. Для полуструктурированных и неструктурированных данных (например, журналы событий, тексты, изображения, видео) часто используются NoSQL базы данных, такие как MongoDB, Cassandra или HBase. Выбор конкретной NoSQL базы данных определяется типом данных и запросами, которые необходимо выполнять. Например, MongoDB хорошо подходит для документов JSON, а Cassandra – для высоконагруженных приложений, требующих высокой доступности.

Apache Spark является одной из наиболее популярных платформ для обработки больших данных. Она позволяет выполнять распределенные вычисления, обрабатывая данные параллельно на множестве узлов кластера, что значительно ускоряет обработку. Spark поддерживает различные языки программирования, такие как Java, Python, Scala и R, что делает его гибким инструментом для решения разнообразных задач. В дополнение к Spark, Hadoop MapReduce также может использоваться для обработки больших данных, хотя Spark обычно превосходит его по скорости выполнения. Выбор между Spark и Hadoop зависит от специфики задач и требований к производительности. Важно отметить, что обработка Big Data часто включает в себя этапы очистки данных (удаление дубликатов, обработка пропущенных значений, коррекция ошибок), трансформации данных (преобразование форматов, агрегация данных) и обогащения данных (добавление новых атрибутов на основе внешних источников). Эти этапы часто требуют применения специализированных инструментов и библиотек, таких как Apache Kafka для обработки потоковых данных или библиотеки машинного обучения, такие как scikit-learn или TensorFlow.

После обработки данные готовы к анализу. Для этого используются различные инструменты, включая специализированные системы бизнес-аналитики (BI), такие как Power BI, Tableau, Qlik Sense. Эти системы позволяют создавать интерактивные отчеты, дашборды и визуализации, облегчая понимание и интерпретацию данных. Они подключаются к различным источникам данных, включая хранилища данных, базы данных и облачные сервисы. Помимо BI-систем, существуют и другие инструменты анализа данных, например, R и Python с соответствующими библиотеками (pandas, dplyr, ggplot2). Выбор инструментов зависит от уровня технической подготовки аналитика и специфики задач.

Безопасность больших данных является критическим аспектом. Необходимо обеспечить конфиденциальность, целостность и доступность данных. Методы обеспечения безопасности включают шифрование данных как в состоянии покоя (на диске), так и в состоянии передачи (по сети), аутентификацию пользователей с использованием многофакторной аутентификации и авторизацию доступа на основе ролей (RBAC), контроль доступа к данным с использованием списков управления доступом (ACL), регулярное резервное копирование данных и мониторинг безопасности. Соблюдение нормативных требований к защите данных, таких как GDPR или CCPA, также является обязательным. Для защиты от несанкционированного доступа и атак необходимо использовать средства обнаружения вторжений (IDS) и предотвращения вторжений (IPS).

Обработка данных в реальном времени. Многие приложения Big Data требуют обработки данных в режиме реального времени. Например, системы мониторинга социальных сетей должны мгновенно анализировать новые сообщения и выявлять тренды. Для этого используются потоковые платформы обработки данных, такие как Apache Kafka, Apache Flink, Apache Storm. Эти платформы позволяют обрабатывать данные по мере их поступления, обеспечивая мгновенную реакцию на изменения. В сочетании с системами машинного обучения, обработка данных в реальном времени может использоваться для создания интеллектуальных систем, способных принимать решения на основе текущей информации.

В заключение, эффективная работа с Big Data требует интегрированного подхода, включающего выбор подходящих инструментов

для хранения, обработки, анализа и визуализации данных, а также надежной системы безопасности. Выбор конкретных технологий зависит от конкретных требований проекта, объема данных, типов данных, необходимой скорости обработки и требований к безопасности. Важно помнить, что это динамично развивающаяся область, и постоянное изучение новых технологий и методов является ключом к успеху.

ЛИТЕРАТУРА

1. Роман Савченко, BigData: особенности обработки и использования больших данных в маркетинге // НРФ 8 2024. – URL: advertisingforum.ru/blog/big-data-v-marketinge-kak-rabotat-s-bolshimi-massivami-dannyh/#block-5.

2. Симакина М.А. Особенности использования технологий в маркетинге // Бюллетень науки и практики. 2018 №6. – URL: cyberleninka.ru/article/n/osobennosti-ispolzovaniya-tehnologiy-big-data-v-marketinge/viewer.

3. Фролова, Д.А., Белая, В.В, Использование Big Data для оптимизации маркетинговых решений // БГУИР 2021. – URL: libelodoc.bsuir.by/handle/123456789/43809.

4. Людмила Клейменова, Ася Зуйкова, “Что такое BigData и почему их называют новой нефтью” // РБК 2023. – URL: trends.rbc.ru/trends/innovation/5d6c020b9a7947a740fea65c.

REFERENCES

1. Roman Savchenko, BigData: Features of Processing and Using Big Data in Marketing // NRF 8 2024. – URL: advertisingforum.ru/blog/big-data-v-marketinge-kak-rabotat-s-bolshimi-massivami-dannyh/#block-5.

2. Simakina M.A. FEATURES OF USING TECHNOLOGIES IN MARKETING // Bulletin of Science and Practice. 2018 No. 6. – URL: cyberleninka.ru/article/n/osobennosti-ispolzovaniya-tehnologiy-big-data-v-marketinge/viewer.

3. Frolova, D.A., Belaya, V.V., Using Big Data to optimize marketing decisions // BSUIR 2021. – URL: ["libelodoc.bsuir.by/handle/123456789/43809"](http://libelodoc.bsuir.by/handle/123456789/43809)4.

4. Lyudmila Kleymenova, Asya Zuikova, “What is Big Data and why are they called the new oil” // RBC 2023. – URL: trends.rbc.ru/trends/innovation/5d6c020b9a7947a740fea65c.