

ИСПОЛЬЗОВАНИЕ БОЛЬШИХ ДАННЫХ В МАРКЕТИНГОВЫХ ИССЛЕДОВАНИЯХ

Большие данные (*Big Data*) – общее название для структурированных и неструктурированных данных огромных объемов. Технологии *Big Data* – это совокупность инструментов, подходов и методов обработки больших данных для дальнейшего их использования. Главной целью является возможность получения на их основе аналитических выводов для принятия решений, поиска новых зависимостей, прогнозирования, достижения бизнес-целей путем анализа, интерпретации и исследования больших объемов данных в условиях непрерывного роста и распределения данных по многочисленным узлам вычислительной сети.

Базовые характеристики *Big Data*: *Volume* (объем), *Velocity* (скорость накопления и обработки постоянно обновляющихся данных), *Variety* (разнообразие типов данных); *Variability* (изменчивость), *Value* (значимость), *Visualization* (визуализация результатов анализа), *Veracity* (достоверность).

Основными задачами *Big Data* являются: построение моделей, основанных на поиске причинно-следственных связей; оптимизация рутинных процессов; составление прогнозов.

Главными особенностями аналитики больших данных является то, что весь объем данных обрабатывается сразу в режиме реального времени, не производится предварительная сортировка данных.

Процесс работы с большими данными можно разделить на следующие три этапа, приведенные в таблице.

Таблица – Этапы процесса работы с большими данными

Этап	Обработка
Сбор данных различного формата из доступных источников:	– интернет вещей (<i>IoT</i>) и другие устройства; – данные от компаний; – региональные и государственные статистические данные; – социальные сети; – <i>GPS</i>-трекеры; – поисковые системы и т. д.
Способ хранения и размещение в хранилище:	– собственные сервера и инфраструктура; – специальные дата-центры; – <i>Data cloud</i> – готовые облачные хранилища.
Анализ и обработка результатов:	– извлечение данных из различных источников, их фильтрация и распределение между узлами; – параллельная обработка на узлах.

Принципы работы с большими данными: расширяемость, отказоустойчивость, локализация обработки в месте хранения.

Выделяют четыре основных метода анализа *Big Data*.

- *Descriptive analytics* (описательная аналитика) – выявление причинно-следственных связей событий, корректировка действий для дальнейшего анализа и принятия решений;

- *Predictive analytic* (прогнозирующая аналитика) – метод прогнозирования на основе оценки текущей ситуации;

- *Prescriptive analytics* (предиктивная аналитика) – выявление и учет проблемных точек в прогнозировании результатов.

- *Diagnostic analytics* (диагностическая аналитика).

Наиболее распространённые базовые технологии обработки больших данных: *NoSQL*, *MapReduce*, *Hadoop*, *R*. Их дополняют аппаратно-программные комплексы, использующие готовые технологические решения, требующие небольших дополнительных настроек или незначительного конфигурирования системы.

В рамках любой маркетинговой компании большие данные помогают проводить анализ в трёх аспектах: общая картина бизнеса, оценка конкурентов, изучение клиентов и целевой аудитории.

Инструменты и технологии для работы с большими данными [1]:

- краудсорсинг – ручной анализ и категоризация источников;

- визуализация – построение графиков, анимированных;

- машинное обучение;

- использование искусственного интеллекта и нейросети;

- смешение и интеграция данных;

- *Data Mining*.

Для хранения и обработки данных применяют технологии:

- классическая реляционная система управления базами данных (СУБД);

- горизонтально масштабируемое хранилище данных (распределённая система);

- система работы с данными *in-memory*.

В системе с горизонтально масштабируемым хранилищем данные распределяются между серверами. Анализ данных происходит на всех серверах параллельно, результат параллельных вычислений консолидируется (*Hadoop*, технология *MapReduce*).

Системы работы с данными *in-memory* позволяют создавать и анализировать структурированные и слабоструктурированные данные в режиме реального времени. Обработка и анализ данных включает в

себя разработку алгоритмов и методов для структурирования, анализа, извлечения информации и получения результатов в реальном времени [2]. Можно выделить следующие инструменты обработки *Big Data*:

Apache Hadoop – фреймворк для обработки больших данных (включает в себя распределённую файловую систему *HDFS* и фреймворк для обработки данных *MapReduce*);

Apache Spark – фреймворк для параллельной обработки данных, который предоставляет *API* для *Java*, *Scala*, *Python* и *R*;

NoSQL-базы данных – масштабируемые и гибкие решения для хранения и обработки неструктурированных данных (*Apache Cassandra*, *MongoDB* и *Couchbase*);

фреймворки для обработки потоков данных в режиме реального времени (*Apache Kafka*, *Apache Storm* и *Apache Flink*);

машинное обучение и искусственный интеллект;

кластерные системы управления данными (предоставляют распределённые хранилища такие, как *Apache HBase*;

технологии для визуализации данных (*Tableau*, *Power BI* и *D3.js*).

Важной задачей в контексте *Big Data* является обеспечение безопасности данных и защита их от несанкционированного доступа.

Эффективная масштабируемость и производительность системы достигается через оптимизацию запросов, распараллеливание обработки данных и использование распределённых вычислений.

Большие данные могут использоваться для обнаружения скрытых закономерностей и трендов. Решение подобных задач заключается в разработке и применении методов машинного обучения и искусственного интеллекта для автоматизации обработки и анализа данных, создания прогностических моделей и принятия интеллектуальных решений.

Примеры применения *Big Data* в маркетинге.

1. Сегментация клиентов методом кластеризации.
2. Создание портрета целевой аудитории.
3. Повышение лояльности клиентов.
4. Визуализация данных и инфографика.
5. Разработка новых продуктов и услуг на основе предиктивного анализа.
6. Снижение коэффициента оттока клиентов.
7. Прогнозирование продаж на основе клиентских метрик.
8. Сокращение издержек (возможность создания карты пути покупателя для разных сегментов аудитории и предсказать наиболее предпочтительные каналы взаимодействия).

В *Big Data* данные хранятся в обезличенном и зашифрованом виде для обеспечения безопасности пользователей.

Обработка данных ведется по принципу от простого к сложному. Вначале данные анализируются и из первоначальной выборки выделяется часть данных, к которым можно применить более простые алгоритмы и методы. Оставшиеся данные передаются на следующий этап обработки, где используются более сложные алгоритмы, и так далее. На последнем этапе сценария обработки применяются сложные алгоритмы к значительно меньшему объему данных. В результате общее время, необходимое для обработки всех данных, уменьшается.

Еще одной эффективной стратегией обработки больших наборов является разбиение данных на сегменты и построение моделей для каждого подмножества с дальнейшим объединением результатов. При этом повышается скорость анализа и снижаются требования к памяти благодаря обработке меньших объемов данных за один проход. Обработка может вестись параллельно и модели для каждого выделенного сегмента данных могут быть уникальными.

Для построения модели на больших объемах исходных данных можно использовать не всю информацию, а некоторое подмножество – репрезентативную выборку. Модель, построенная на относительно небольшом множестве, применяется в дальнейшем ко всему набору данных, соответственно, время получения результата сократится по сравнению с попыткой обучения модели на всем имеющемся объеме данных.

Маркетинговые исследования – это прежде всего математический расчет. Маркетинговые исследования всегда сопряжены анализом большого объема информации. Технологии *Big Data* являются инструментом построения бизнес-стратегий и маркетинговых исследований для достижения компаниями бизнес-целей и позволяют создавать новые формы онлайн- или цифрового маркетинга.

ЛИТЕРАТУРА

1. Большие данные (Big Data) [Электронный ресурс] – Режим доступа:

[https://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F:%D0%91%D0%BE%D0%BB%D1%8C%D1%88%D0%B8%D0%B5_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5_\(Big_Data\)](https://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F:%D0%91%D0%BE%D0%BB%D1%8C%D1%88%D0%B8%D0%B5_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5_(Big_Data)) – Дата обращения: 20.01.2024.

2. Анналин Б.Н. Теоретический минимум по Big Data. Всё, что нужно знать о больших данных. / Б.Н. Анналин, С.У. Кеннет. – СПб.: Питер, 2019. – 208 с.