

УДК 33.338

**«БОЛЬШИЕ ДАННЫЕ» КАК ИНСТРУМЕНТ  
СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ  
В СОВРЕМЕННОМ МИРЕ**

**Канатова А.М.**

Евразийский Национальный Университет имени Л.Н. Гумилёва  
г. Нур-Султан, Казахстан

Научный руководитель – **Шалболова У.Ж.**

Последние годы в развитии современного мира были полны фундаментальных научно-технических инноваций. Они предоставили человечеству набор новых возможностей для сбора, хранения и анализа цифровых данных. Количество данных и их источников в современном мире растет в геометрической прогрессии. Каждое действие в Интернете может быть документировано, обработано и подключено к большому количеству других действий и данных. В результате стремительно накапливается новая информация, ее объем растет. Это позволяет использовать информационные ресурсы в различных областях и для различных целей. Таким образом, вопрос технологии обработки данных очень актуален сегодня. В первую очередь это касается такой инновационной технологии, как «большие данные», которой отдают предпочтение многие современные пользователи.

«Большие данные» является эффективным инструментом для обработки больших объемов информации и принятия на ее основе индивидуальных решений. Это очень полезно для современных людей. Большие объемы данных обрабатываются таким образом, что каждый может получить конкретные и желаемые результаты для их дальнейшего эффективного использования. Понятие "большие данные" подразумевает наборы данных, которые являются объемными, сложными, плохо структурированными и быстро меняющимися [5]. Таким образом, они не могут быть обработаны традиционным способом. Обычно это более 100 ТБ данных в день [1]. "Большой объем данных" иллюстрирует комплекс технологий, функционирующих на основе обработки данных и цифровых коммуникаций. Возможно, в будущем эти технологии вызовут основные изменения в глобальном мире и обществе.

«Большие данные» являются одной из ведущих формирующихся технологий, которая в ближайшие десятилетия внесет большой вклад в различные области науки и окажет значительное воздействие на различные аспекты человеческого общества" [11]. Сегодня большие объемы данных уже генерируются во всех сферах человеческой деятельности. Мы оперируем ими повсюду. Спам-фильтры в нашей электронной почте автоматически адаптируются к различным вирусным угрозам и нежелательным письмам. Автомобильный трафик успешно прогнозируется на несколько часов вперед с помощью специального спутникового слежения. Анализ фактов и данных, не имеющих связи с первого взгляда, демонстрирует нам удивительные корреляции в различных сферах жизни и деятельности.

Сегодня невозможно представить современный мир без технологий «большие данные». Они полезны во всех сферах человеческой деятельности - от медицины и общественного порядка до маркетинга и продаж. Данные постоянно создаются и собираются государственными организациями и частными компаниями. Затем они обрабатываются с помощью цифровых технологий. Такая информация ежедневно обновляется и расширяется. Она включает в себя различные факты из области систем связи, измерительных и контрольных приборов, секторов экономики, страхования и банковского дела, планирования и диагностики, статистических отчетов, медицинской помощи, научных исследований и образования, издательской деятельности в новостных лентах и социальных сетях и др.

Одним из важнейших компонентов «больших данных» являются «социальные данные». Социальные данные — это часть «больших данных», создаваемых людьми для некоммерческих целей, в том числе различных социальных сетей, фотобанков, блогов, чатов и т.д." [1]. Миллионы пользователей ежедневно посещают популярные веб-сайты и платформы и оставляют после себя различные "следы" [2]. Они пишут комментарии, публикуют обзоры и обмениваются мнениями. Такая информация и другие частные данные доступны для анализа и поиска корреляций. Это уникальный ресурс, который можно сравнить с "нефтью" XXI века. Его также называют настоящей "революцией социальных данных" [10], основанной на растущем влиянии Интернета на жизнь и деятельность людей.

«Революция социальных данных» — это популярная в настоящее время тенденция изменения образа жизни человека и его взаимодействий [3]. Она особенно заметна в последнее десятилетие. Она проявляется в виде обмена личной информацией и его различных последствий. В основе этих тенденций лежит массовое распространение социальных сетей и их возрастающее влияние на различные сферы социальной жизни. Беспрецедентное накопление опубликованных частных данных является феноменом современности. Это уникальный источник информации, который постоянно обновляется. Его можно эффективно использовать в социальных науках и цифровых гуманитарных науках для изучения человеческого поведения - индивидуального или коллективного, на уровне социальных групп или общества в целом.

В маркетинге инструменты «большие данные» позволяют выявить наиболее эффективные идеи на определенном этапе цикла продаж. Анализ данных используется для улучшения отношений между инвестициями и системой управления клиентами, выбора эффективной стратегии увеличения коэффициента конверсии и оптимизации жизненного цикла клиента. В бизнесе, связанном с облачными технологиями, алгоритмы «большие данные» используются для минимизации затрат на привлечение клиента и увеличения его жизненного цикла.

Основной причиной использования «большие данные» в маркетинге является необходимость дифференциации ценовых стратегий. Последняя зависит от индивидуальных привычек клиента и уровня его внутренней системы. Глобальный институт МакКинси установил, что около 75% доходов среднестатистической фирмы поступает от основных продуктов, но 30% из них оцениваются неправильно. Рост цен на 1% приводит к увеличению операционной прибыли на 8,7% [8]. Исследовательская группа Форрестер обнаружила, что анализ данных позволяет маркетологам сделать отношения с клиентами более успешными. Исследования в области развития клиентов помогают специалистам оценить уровень их лояльности, а также продлить жизненный цикл в рамках определенной компании [9].

Анализ данных помогает современным компаниям получить полное представление о ключевых аспектах их бизнеса. Увеличение доходов, снижение затрат и оборотных средств — вот три основные задачи, которые современный бизнес пытается решить с помощью аналитических инструментов.

Для изучения и прогнозирования поведения клиентов обычно используются следующие методы: прогнозирование (классификация, регрессионный анализ и т.д.) - для прогнозирования конкретного показателя, например, продукта, который выберет покупатель; сегментация (кластеризация) - разделение людей на группы со схожим поведением. В дальнейшем необходимо работать с каждой группой в отдельности, исходя из схожих потребностей и интересов ее клиентов. Например, если вы выберете группу с высокой закупочной ценой, вы сможете предложить им самые дорогие продукты, правила ассоциации - находить типичные модели поведения покупателей, позволяющие давать рекомендации по продуктам/группам продуктов. Например, клиент выбирает пульт дистанционного управления, и

система предлагает ему батареи автоматически. Этот метод часто используется для интернет-магазинов, нейронные сети - обучающие алгоритмы, позволяющие сегментировать клиентов и предсказывать их поведение.

Использование этих методов дает новые данные, улучшающие модели во всех областях. С маркетинговой точки зрения, новая информация о продуктах и их использовании позволит создать новую модель взаимодействия с клиентами. В частности, будущие предложения будут полностью персонализированы. Анализ новых данных даст новые знания, полезные для рекламных компаний и платформ. Это поможет предсказать поведение клиента и предложить ему то, что ему действительно нужно, или то, о чем у него нет времени даже думать.

Использование технологий «большие данные» вызывает ряд проблем, которые уже обсуждались в исследовательской литературе. Например, многие авторы отмечают "неизбежность постоянного повышения роли данных в нашем обществе" [6], описывают "общество, основанное на данных, в котором данные стали одним из самых ценных активов" [11]. Таким образом, современным пользователям и даже специалистам очень сложно обрабатывать растущий объем информации. Традиционный подход к этой проблеме, связанный с попыткой структурирования баз данных, неэффективен по отношению к огромным информационным ресурсам. Единственная технология, которая действительно может помочь нам в такой проблемной ситуации — это "большие данные". Она может обрабатывать плохо структурированные данные, конкретную информацию, хранящуюся в различных форматах и периодически обновляемую.

Вопрос о качестве неструктурированных данных также очень актуален сегодня. Это связано с большим количеством фальсифицированного контента, находящегося в Интернете. Например, существуют специальные агенты и программы искусственного интеллекта для написания обзоров как позитивного, так и негативного контента. Дело в том, что технология «большие данные» ежедневно обрабатывает большие объемы информации [4]. Поэтому требования к ее точности должны быть снижены. Когда возможности измерения ограничены, мы учитываем важнейшие показатели только по логическому желанию получить точную цифру. Известно, что тщательная проверка информации невозможна без требования точности. Если говорить о небольших объемах информации и подобных случаях, то эта процедура выглядит достаточно важной. Но если мы имеем в виду технологию «большие данные», то точность кажется невозможной и даже ненужной, особенно в ее абсолютном виде.

Если вы работаете с постоянно меняющимися данными, то строгая точность занимает особое значение. Большие данные полностью не упорядочены, в их состав входит огромное количество ресурсов различного качества и природы, а бесчисленные серверы расположены по всему миру [3]. На уровне больших данных мы обычно формируем общий взгляд на предмет, его тщательное изучение со всеми нюансами и деталями не требуется. Если мы теряем часть информации из-за ее неаккуратности на микроуровне, то это позволяет нам делать открытия на макроуровне. Мы теряем точность, но можем

обнаружить внезапные корреляции между данными различной природы. А это обеспечивает рост знаний, открывает ряд новых перспектив и возможностей.

Многие авторы обращают внимание на следующее обстоятельство: неправильно собирать как можно больше данных, когда работаешь с «большими данными» [7]. Сбор данных ради самих данных, а также энтузиазм нескольких специалистов к новым технологиям ради этих технологий, непродуктивен. Нужно учитывать специфику реальной практики.

Одним из самых известных и широко известных применений инструментов «большие данные» является обработка запросов пользователей и таргетинг. «Большие данные» помогает нам анализировать привычки клиентов, чтобы в будущем понять их потенциальные запросы. Современные компании используют не только традиционные наборы данных и информационные ресурсы. Последние расширяются за счет информации из социальных сетей («социальные данные»). Изучение истории поиска браузера также очень эффективно: это помогает сформировать наиболее полную клиентскую базу и получить полную информацию о ее участниках. Если руководитель компании хочет двигаться в ногу со временем и успешно развиваться, ему необходимо создать собственную прогнозную модель с помощью технологии «большие данные». К сожалению, на сегодняшний день к этому готовы только крупные компании. В целом, объем данных, обрабатываемых технологией «большие данные», постоянно растет, как и скорость их обработки. Развитие этого направления вполне созвучно современному миру, его стремительности и инновациям.

### Список литературы:

1. Денисова О.Ю., Мухутдинов Е.А.: Большие данные - Речь идет не только о размере данных. Технологический университет 18(4), 226-230 (2015).
2. Гурьянова А., Астафьева Н., Филатова Н., Хафиятуллина Е., Гурьянов Н. Философские проблемы информационно-коммуникационных технологий в процессе современного социально-экономического развития. В: Попкова Е.Г., Островская В.Н. (ред.) Перспективы использования новых информационно-коммуникационных технологий (ИКТ) в современной экономике. Достижения интеллектуальных систем и вычислительной техники, т. 726, с. 1033-1040. Спрингер, Чам (2019 год)
3. Гурьянова А.В., Хафиятуллина Е., Петина М., Фролов В., Маховиков А. Технологические предпосылки и гуманитарные последствия повсеместных вычислений и сетевого взаимодействия. Ин: Попкова Е.Г., Сергей Б.С. (ред.) Цифровая экономика: Сложность и многообразие против рациональности. Лекционные заметки по сетям и системам, т. 87, с. 1040-1047. Спрингер, Чам (2020 г.)
4. Япкович, Н., Стефановски, Я. (ред.) Анализ больших данных: Новые алгоритмы для нового общества, Спрингер, Чам (2016).
5. Климович А.: Вопросы философии больших данных. Инновации. Ученые А. Ж. 8(84), 30-38 (2018).

6. Клоус, С., Вилаард, Н.: Мы - Большие Данные. Будущее информационного общества, Атлантис Пресс, Париж (2016).
7. Майер-Шенбергер, В., Кукиер, К.: Большие данные: Революция, которая изменит то, как мы живем, работаем и думаем. Хаутон Миффлин Харкорт, Нью-Йорк (2013).
8. Глобальный институт МакКинси: Большие данные: Следующая граница инноваций, конкуренции и производительности (2011). <https://lazowska.cs.washington.edu/escience/McKinsey.big.data.pdf>. Доступно 21 июня 2020 г.
9. Аналитический обзор рынка Больших данных (2015). <https://habr.com/ru/company/moex/blog/256747/>. Доступно: 21 июня 2020.
10. Вейгенд, А.: Данные для людей: как заставить нашу экономику пост-привилегии работать на вас. Базовые книги, Нью-Йорк (2017 г.)
11. Зомая А.Ю., Сакр Ш. (ред.) Энциклопедия технологий больших данных. Спрингер, (2019)