

УДК 004.9:339.138

ИСПОЛЬЗОВАНИЕ DATA SCIENCE В ЭЛЕКТРОННОМ МАРКЕТИНГЕ: СТРУКТУРА, МЕТОДЫ И ПОКАЗАТЕЛИ ЭФФЕКТИВНОСТИ



А.Д. Погорецкая

Студент 4 курса специальности "Электронный маркетинг" инженерно-экономического факультета БГУИР



О.Н. Шкор

Старший преподаватель, магистр экономических наук, доктор философии в области экономики

Белорусский государственный университет информатики и радиоэлектроники

Г. Минск, Республика Беларусь

E-mail: shkor@bsuir.by, a.pogoretskaya@gmail.com

О.Н. Шкор

Родилась в Минске. Закончила БПИ в 1984 году. В 2001 г. защитила магистерскую диссертацию на тему: «Использование блочно-модульной системы обучения в профессиональной ориентации школьников» по специальности 08.00.05 «Экономика и управление народным хозяйством». В 2010 г. защитила докторскую диссертацию (PhD) на тему: «Экономическое обоснование формирования транспортной логистической системы Республики Беларусь» в Международной кадровой академии (Киев). С 2014 г. по настоящее время - заместитель заведующего кафедрой по научно-исследовательской работе студентов.

А.Д. Погорецкая

Родилась в 2000 году в Минске. В 2018 году закончила ГУО «Средняя школа №4 г. Минска». В этом же году поступила в УО «БГУИР», была зачислена на платную форму обучения по специальности «Электронный маркетинг» инженерно-экономического факультета.

Аннотация. Использование Data Science за последние 10 лет значительно увеличилось. Эта технология облегчает принятие решений и извлечение практических знаний из больших объемов информации в среде электронного маркетинга. Несмотря на эти достижения, соответствующих фактических данных о мерах по улучшению управления технологией Data Science в электронном маркетинге недостаточно. В данной статье будут рассмотрены методы анализа, использования и показатели эффективности Data Science, которые используются в методах и стратегиях электронного маркетинга. Результаты исследования представляют собой целостный обзор основных технологий Data Sciences в электронном маркетинге и дают представление, связанное с созданием инновационных методов интеллектуального анализа данных и обнаружения знаний.

Ключевые слова: Data Science, Big Data, электронный маркетинг, исследование литературы.

Введение. Термин Big Data возник во 2008 г.. В Первый Раз его применил редактор журнала Nature — Клиффорд Линч. Он рассказывал про увеличение объемов международных данных также отмечал, то что овладеть ими могут помочь новейшие технологии.

С начала 21 века электронный маркетинг и Data Science заметно эволюционировали с точки зрения использования и прибыльности [1]. Это привело к появлению цифровой экосистемы, которая связывает пользователей 24/7, это в свою очередь сформировало у пользователей новые привычки и поведение. Электронный маркетинг определяется как набор методов, разработанных в Интернете для того, чтобы убедить пользователей купить продукт или услугу. На сегодняшний день, компании, которые работают в Интернете используют такие

методы, как поисковая оптимизация (SEO), т.е. оптимизация результатов поиска из основных поисковых систем; Маркетинг в поисковых системах (SEM) или программная реклама, то есть стратегии спонсирования рекламы в поисковых системах или в рекламных местах на баннерах на веб-сайтах; а также маркетинг в социальных сетях (SMM), то есть стратегии взаимодействия с пользователями в социальных сетях посредством социальной рекламы [2].

В последнее десятилетие электронный маркетинг вызвал значительный исследовательский интерес среди ученых [9]. Например, Роджерс и Секстон (2012) стремились понять ключевые способы повышения прибыльности или ROI (возврат инвестиций) в электронном маркетинге. Кроме того, Кумар (2013) измерил влияние данных на экосистему электронного маркетинга. Аналогичным образом Саура, Палос-Санчес и Серда Суарес (2017) определили метрики для измерения эффективности каждого из действий электронного маркетинга, которые использовала компания в Интернете.

Многочисленные исследования показали, что ключевой способ увеличения эффективности стратегии электронного маркетинга заключается в применении методов Data Science в этой отрасли [3]. Например, Келлехер и Тирни (2018) утверждали, что Data Science может повысить эффективность электронного маркетинга за счет улучшения таких вопросов, как управление компаниями информацией, полученной от пользователей; тип и источник данных из наборов данных компаний, и применение нового анализа данных и инновационных методов для создания знаний [4].

Кроме того, Фань, Лау и Чжао (2015) и Саура и Беннетт (2019) подчеркнули важность нескольких аспектов, такие как тип данных, собранных из различных онлайн-источников, покупки, сделанные пользователями, и их поведение в Интернете.

Аналогичным образом Ведель и Каннан (2016) продемонстрировали, что для увеличения шансов на успех в социальных сетях, компании должны выявлять неожиданные закономерности, используя методы искусственного интеллекта (AI) или машинного обучения (ML).

Соответственно, индустрия электронного маркетинга находится под большим влиянием таких областей исследований, как информационные науки (IS) или компьютерные науки (CS), а также всеми другими областями исследований, которые облегчают сбор, упорядочивание и управление данными.

В целом, основной целью Data Science является получение знаний из анализа данных, чтобы ответить на конкретные исследовательские вопросы [5]. Анализируя данные, методы Data Science позволяют извлекать шаблоны из баз данных, чтобы объяснить проблему или сформулировать поздние гипотезы. В Data Science ключевая идея заключается в том, что шаблоны, выявленные в данные не очевидны и полезны для компаний. С точки зрения обнаружения закономерностей люди могут идентифицировать характеристики предмета (товар, услуга, сообщество и т. д.).

Шаблоны, выявленные с помощью методов Data Science, помогают получить полезную информацию, то есть то, что исследователи или специалисты по данным хотят извлечь из выявленных паттернов [5]. Типы шаблонов в Data Science: *кластеры* (сегментация в бизнесе, которая состоит из идентификации поведения, вкусов или привычек, которые выявляют одинаковую группу потребителей), *обучение ассоциативным правилам (ARM)* (заключается в выявлении закономерностей: продукты, купленные одновременно или мошенничество в финансовом или страховом секторах), модели прогнозирования (шаблон, предсказывающий пропущенное значение атрибута (товар, услуга и т.д)).

Шаблоны, выявленные с помощью методов Data Science, помогают получить полезную информацию, то есть то, что специалисты по Data Science хотят извлечь из выявленных паттернов [5].

Big Data характеризуется: объемом, т. е. чрезмерном количеством данных, разнообразием типов данных и скоростью, с которой данные должны обрабатываться [5]. Основа Big Data в том виде, в каком они существуют сегодня, была заложена Моделью реляционных данных (RDM) Кодда (1971), которая позволяла чтение и хранение информации, а также выполнение прямых

запросов по информации в базах данных. физическое расположение базы данных. Это усовершенствование устранило проблему физического расположения базы данных. Это была важная веха в индустрии Data Science — раньше базы данных находились в отдельных физическом хранилищах [6].

Кодд (1971) также заложил основы языка структурированных запросов (SQL), современного стандарта запросов к базам данных. Последние разработки в области хранения данных привели к появлению новых баз данных NoSQL. Они хранят различные данные и их атрибуты с такими языками, как JavaScript Object Notation (JSON). JSON меньше весит, имеет более высокий процесс-скорость и самоописание, а не на основе таблицы реляционной модели, такая как SQL.

Происхождение и источники данных также важны в Data Science. В зависимости от происхождения данных могут использоваться различные типы Data Science. Типы источников данных и описания в Data Science в электронном маркетинге: *транзакционные данные* (информация о продажах, счета-фактуры, квитанции, отгрузки, платежи, страхование, аренда); *нетранзакционные данные* (демографические, психографические, поведенческие, данные об образе жизни); *эксплуатационные данные* (данные о стратегиях и действиях, связанных с логистика и бизнес-операции); *онлайн-данные* (пользовательский контент, электронные письма, фотографии, твиты, лайки, акции, веб-сайты, веб-поиск, видео, онлайн-покупки, музыка).

В Data Science базы данных состоят из различных переменных или индикаторов. Эти базы данных известны как «наборы данных» или «записи данных». Каждая из переменных содержащаяся в наборах данных, обозначает конкретную характеристику. Наборы данных могут содержать структурированные или неструктурированные данные. Структурированные данные могут храниться в таблицах, каждая из которых имеет одинаковую структуру или атрибуты. Неструктурированные данные имеют свою внутреннюю структуру и, следовательно, атрибуты могут быть организованы по-разному в каждой таблице.

Для анализа набора данных Data Science полагается на модели, основанные на машинном обучении (ML). Оно предоставляет алгоритмы для автоматического анализа больших наборов данных. Эти модели могут обучать исследователи или маркетологи для извлечения полезной информации и выявления закономерностей.

ML превратился в технологию глубокое изучение (DL). Она позволяет нам изменить то, как компьютеры обрабатывают языки и образы. DL состоит из набора моделей нейронных сетей с несколькими слоями и модулями в одной сети [7]. Существуют и другие подходы к данным с использованием ML.

В области ML существует два основных типа подходов к анализу: контролируемое обучение (SL) и неконтролируемое обучение (UL). SL включает в себя обучение набора образцов, включая фрагменты текста, пользовательский контент (UGC), например твиты или посты в Facebook и так далее. Все эти примеры можно использовать для обучения алгоритма.

Data Science предлагает различные точки зрения и подходы к статистическому анализу данных. Статистика — это набор правил для количественного анализа любого типа данных. Однако с развитием математики и Data Science, статистическое обучение было определено как теоретическая основа, которая работает с ML с точки зрения Data Science [8].

Таким образом, цель методы, используемые в Data Science, применяющие статистическое обучение, должны выполнять функциональный анализ, исследовательский анализ и прогнозировать результаты на основе проанализированных данных. Методы, определенные и применяемые в экосистеме DM:

1. Описательная статистика - используется для количественного обобщения признаков из набора информации или набора данных. Она включает измерение центральной тенденции, среднего арифметического или таких показателей, как дисперсия или диапазон.
2. Правило Байера - используется для описания вероятностей событий. Оно основано на знании условий, в которых могло вызвать конкретное событие.

3. Метод наименьших полных квадратов - позволяет найти наилучшую теоретическую модель, состоящую из переменных или конструкций, которые соответствуют набору данных и позволяют проводить количественную проверку.

4. Линейная регрессия - используется для моделирования связи между шкалой и исследовательской переменной.

5. Искусственные нейронные сети - это самообучающиеся системы, состоящие из взаимосвязанных нейронов узлов, которые имеют вход и вывод. Они используются для поиска или обнаружения решений, которые иначе трудно идентифицировать с помощью стандартного программирования.

6. Многофакторный анализ - модель анализа, ориентированная на множественный анализ данных, собранных из более чем одного зависимая переменная.

7. Искусственный интеллект - ориентирован на моделирование человеческого интеллекта машинами. Он относится к использованию машин, автоматически ориентированных на машинное обучение и решение проблем.

В экосистеме электронного маркетинга одной из основной задачи является определение стратегии Data Science. Для этого маркетологи и должны выбрать и понять основные характеристики метрик и показатели эффективности для измерения моделей и методологий.

Показатели эффективности Data Science в электронном маркетинге:

1. Надежность - показатель точности и полноты обработки набора данных, связанный с его использованием.

2. Точность - показатель, который измеряет актуальность и успешность подхода метод в базе данных, где был применен Data Science

3. Положительная прогностическая ценность (PPV) - представляет собой показатель, который измеряет актуальность и успешность подхода метод в базе данных, где был применен Data Science.

4. Согласованность - оценивает, соответствуют ли значения, представленные в данных из одного набора данных, значениям, представленным в данных.

5. Чувствительность - относится к количеству правильных результатов, деленному на количество отброшенных значений.

6. Специфичность - показатель, который оценивает прогностические характеристики истинных отрицательных результатов в переменных в категории в наборе данных.

7. Распространенность - относится к доле населения с определенными общими характеристиками в течение определенного периода времени.

Заключение. В этой статье были описаны основные понятия, методы и показатели эффективности, используемые в Data Science на протяжении последних двух лет. Соответствующие методы, используемые в Data Science для извлечения полезной информации также были идентифицированы из больших объемов данных.

Сегодня компании во все большей степени вовлечены в работу с данными. Соответственно, количество компаний, использующих ML значительно увеличилось. Очень важно иметь квалифицированных специалистов в области электронного маркетинга, которые владеют основными понятиями Data Science, для того чтобы проводить более качественные исследования и структурировать свои базы данных.

Список литературы

[1] Tiago, M. T. P. M. B., & Veríssimo, J. M. C. (2014). Digital marketing and social media: Why bother? *Business horizons*, 57(6), 703–708.

[2] Lies, J. (2019). Marketing intelligence and big data: digital marketing techniques on their way to becoming social engineering techniques in marketing. *International Journal of Interactive Multimedia & Artificial Intelligence*, 5(5).

[3] Braverman, S. (2015). Global review of data-driven marketing and advertising. *Journal of Direct Data and Digital Marketing Practice*, 16(3), 181–183.

[4] Palacios-Marqués, D., García, M. G., Sánchez, M. M., & Mari, M. P. A. (2019). Social entrepreneurship and organizational performance: A study of the mediating role of distinctive competencies in marketing. *Journal of Business Research*, 101, 426–432.

[5] Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT Press.

[6] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends Data Science® in Theoretical Computer Science*, 9(3–4), 211–407.

[7] Lies, J. (2019). Marketing intelligence and big data: digital marketing techniques on their way to becoming social engineering techniques in marketing. *International Journal of Interactive Multimedia & Artificial Intelligence*, 5(5).

[8] Tsapatsoulis, N., & Djouvas, C. (2019). Opinion mining from social media short texts: Does collective intelligence beat deep learning? *Frontiers in Robotics and AI*, 5, 138.

[9] Kannan, P. K. (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, 34(1), 22–45.

USING DATA SCIENCE IN DIGITAL MARKETING: STRUCTURE, METHODS AND PERFORMANCE MEASUREMENTS

O.N. SHKOR

*Senior Lecturer at the
Department of Economics of
BSUIR*

A.D. Pogoretskaya
Student of BSUIR

Belarusian State University of Informatics and Radioelectronics G. Minsk, Republic of Belarus, Senior Lecturer at the Department of Economics, shkor@bsuir.by

Annotation. The use of Data Science has increased significantly over the past 10 years. This concept involves making decisions and extracting practical knowledge from a large amount of information in a wide market. Despite these advances, there is insufficient evidence on measures to improve data science management in conjunction with marketing. This article discusses methods for analyzing, using and evaluating the effectiveness of Data Science, which are used in effective marketing methods and strategies. The results of the study, carried out in conjunction with an overview of the main technologies of Data Sciences in e-marketing, offer insights related to the creation of methods for discovery, data mining and knowledge retrieval.

Keywords: Data Science, Big Data, digital marketing, literature research.