

*Струнин А.Е.
студент
Ярославский Государственный университет им. П.Г. Демидова
Россия, г. Ярославль
e-mail: aleksey@strounine.com*

ИССЛЕДОВАНИЯ ПО АНАЛИЗУ БОЛЬШИХ ДАННЫХ

***Аннотация:** В статье приводится обзор исследований по большим данным (big data) их непосредственном использовании в бизнесе, медицине и т.д.*

Ключевые слова: большие данные, оптимизация бизнеса, технологические инновации, проблема безопасности, инновационные приложения.

*Strunin A.E.
student
Yaroslavl State University named after P.G. Demidova
Russia, Yaroslavl*

BIG DATA ANALYSIS RESEARCH

***Abstract:** The article provides an overview of research on big data (big data) and its direct use in business, medicine and so on.*

Keywords: big data, business optimization, technological innovations innovation, security issues, innovative applications.

Анализ больших данных не является чрезвычайно новым явлением в мире, потому что он существовал и обсуждался многими аналитиками и программистами больше десяти лет назад [1]. Анализ больших данных - это обработка большого количества информации для различных целей, например, для определения потребительских предпочтений и рыночных тенденций или любая другая информация, связанная, к примеру, с бизнесом. Многие коммерческие структуры пользуются анализом больших данных в принятии решений для поддержания конкурентноспособного бизнеса.

За много лет существования больших данных — многое изменилось. Всё больше информации находится в интернете и благодаря этому можно провести достаточно большую работу по анализу данных из открытых источников для

того, чтобы выяснить спрос на тот или иной товар/поведенческий психотип и т.д.

Эти чрезвычайно большие данные в наше время удобно хранятся в базах данных, доступных любому желающему. Согласно оценке, сделанной компанией IBM, опубликованной недавно, было выпущено значительное количество данных объемом 2,5 млрд Гбайт [2].

Каждый день во всем мире этот объем, очевидно, увеличивается каждую секунду. В современном мире аналитикам постоянно удается справляться с огромным количеством «сырой» информации и чрезвычайно сложными паттернами данных. Это достигается за счет использования высокоэффективных технологий, алгоритмов корреляции статистики, хранения инфраструктуры о ценообразовании и экстенсивных методах в добыче полезных ископаемых. Примеры источников больших данных разнообразны и богаты по своей природе, начиная от крупного корпоративного бизнеса, так и правительственные инфраструктуры доступные в интернете. Это огромный объем различных запросов, записи пользователей и все их онлайн-взаимодействия в социальных сетях.

Применение больших данных является ключевым моментом в современном инновационном бизнесе. Благодаря этому анализу в коммерческой инфраструктуре можно провести аналитику по актуальности того или иного товара или какой-либо услуге на рынке [3].

Так же можно сделать конкретные выводы, какие требования ставит перед собой потребитель относительно той или иной продукции, а также какие есть отторгающие факторы для конкретного клиента.

Всё это нужно для того, чтобы максимально привлечь потребителя к себе посредством разных маркетинговых уловок, рассчитанных под конкретного человека (скидки, промо-акции и т.д.).

Помимо этого, можно провести аналитику по актуальности того или иного товара или услуге на рынке, и оценить существующую конкуренцию в данный момент.

Если делать конкретный вывод, то всё это делается для увеличения числа продаж и привлечения большего числа клиентов к конкретной коммерческой организации, для того, чтобы получить из этого максимально возможную выгоду.

Сектор здравоохранения с недавних пор так же начал использовать анализ больших данных. Помимо коммерческого применения, для улучшения продаж каких-либо медицинских препаратов или услуг, применяется аналитика со стороны государственного здравоохранения [4]. Это делают для того, чтобы производить статистический сбор данных, для того, чтобы определить «зоны риска» в плане каких-либо эпидемий или конкретных болезней, для того, чтобы максимально выгодно распределять ресурсы медицинских организаций.

Так же эта аналитика данных в медицинской сфере полезна каждому жителю того или иного региона [5]. Благодаря ей, человек может быть уже готов к каким-либо конкретным вирусам, которые характерны для данного региона проживания.

Список литературы:

1. Hope T., Resheff Y., Lieder I. Learning TensorFlow. Boston: Oreilly, 2017. 242 pp.
2. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge, MA: MIT Press, 2016. 800 pp.
3. Dua D. and Graff C. UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science, 2019.
4. Foster Provost, Tom Fawcett. Data Science for Business. NYU: O'reilly, 2013. 77 pp.
5. Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia. Learning Spark: Lightning-Fast Big Data Analysis. NYU: O'reilly, 2015. 274 pp.