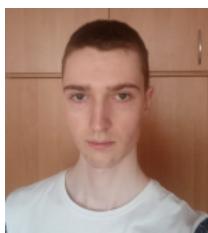


УДК 004.62:316.472.4

ПРИМЕНЕНИЕ BIG DATA ДЛЯ АНАЛИЗА СОЦИАЛЬНЫХ СЕТЕЙ



Е.И. Баяк

Студент специальности ИиТП
факультета компьютерных
систем и сетей БГУИР
ibayak@gmail.com



С.Н. Нестеренков

Декан факультета компьютерных
систем и сетей БГУИР, кандидат
технических наук, доцент
s.nesterenkov@bsuir.by



Д.А. Жалейко

Инженер-программист
ОИТ БГУИР
d.zhalejko@bsuir.by

Е.И. Баяк

Окончил среднюю школу №13 города Гродно. Обучается в Белорусском государственном университете информатики и радиоэлектроники.

С.Н. Нестеренков

Кандидат технических наук, доцент, декан факультета компьютерных систем и сетей Белорусского государственного университета информатики и радиоэлектроники, доцент кафедры программного обеспечения информационных технологий. Автор публикаций на тему машинного обучения, алгоритмов принятия решений, искусственных нейронных сетей и автоматизации.

Д.А. Жалейко

Окончил в Белорусский государственный университет информатики и радиоэлектроники в 2021 году по специальности "Вычислительные машины, системы и сети".

Аннотация. Большие данные или Big Data современный термин, описывающий огромные объемы как структурированной, так и неструктурированной информации. Социальные сети, как источник больших данных, имеют отличительные черты в их сборе и методах анализа. А полученные результаты применяются в различных сферах, начиная от обеспечения информационной безопасности и заканчивая улучшением финансовых показателей.

Ключевые слова: Прогнозная аналитика, информационная безопасность, интеллектуальный анализ данных, социальная сеть.

Введение.

Термин большие данные появился в 2008 году. Big Data – это огромное разнообразие данных, которые постоянно поступают в хранилище в больших объемах. К большим данным относят аббревиатуру VVV (объем, скорость, разнообразие). То есть большие данные характеризуются многообразием, быстрым поступлением и большим объемом данных [1, 2]. Но почему этот термин зачастую связывают с социальными сетями?

По состоянию на январь 2023 года за год количество заходящих в социальные сети людей увеличилось на 137 миллионов человек. Среднее ежедневное использование составляет 2 часа 31 минуту. Большая часть населения мира (4,76 миллиарда человек) пользуются социальными сетями. Следовательно, к источникам больших данных можно отнести социальные сети или мессенджеры. В этом случае Big Data представляет информацию о зарегистрированных пользователях в социальной сети, их информацию, картинки профилей и сообщения которыми обмениваются миллионы людей ежедневно. В некоторых случаях платформа может собирать информацию о кликах и разнообразных действиях пользователей. Различные изображения аудио-, видеофайлы поступают в неструктурированном виде и не могут быть сохранены в таблицах традиционных реляционных баз данных. Поэтому различные мессенджеры являются наглядным источником больших данных, так как информация разнообразна, постоянно изменяется, добавляется и удаляется.

Аспекты применения Big Data для анализа социальных сетей.

Большие данные могут быть применены в различных технологиях. Для социальной сети сбор и анализ информации о действиях пользователей, количестве зарегистрированных, активных за определенный период времени участников является обязательным для улучшения и разработки. Настоящие технологии позволяют получить полезные данные о качестве обслуживания клиентов гораздо эффективнее, чем раньше. С помощью Big Data можно извлечь полезные сведения из социальных сетей, таким образом повысив качество взаимодействия с клиентами и сделав предложения максимально полезными.

Информационная безопасность – это практика предотвращения несанкционированного доступа, раскрытия, использования, искажения, изменения, записи или уничтожения информации в физическом или электронном виде. Её задача – обеспечивать защиту конфиденциальности, целостности и доступности определенных данных. Поиск вредоносной активности обеспечивается путем сбора и анализа огромных объемов данных. Так, в социальных сетях частой проблемой является взлом аккаунта, спам, создание многочисленных так называемых интернет-ботов и различные интернет-мошенничества. Большие данные распознают шаблоны, характерные для мошенников, и собирать значительные объемы данных, чтобы ускорить предоставление нормативной отчетности о безопасности [3].

Big Data также играет роль в финансовой эффективности компании. Среди двух основных способов финансирования выделяют рекламные интеграции и подписку на услуги. Результаты исследований данных повышают эффективность финансовых решений и планирования.

Проведение анализа Big Data.

Существует три способа классификации методов анализа для социальных сетей: по типу данных, по цели и в зависимости от характера задачи. Чаще всего анализируют текст, изображения, видео, аудио. По цели анализа выделяют: предиктивный, описательный, предписывающий и диагностический метод. А в зависимости от характера задач можно проводить визуализацию либо веб-аналитику. Существуют различные методы анализа больших данных, такие как: интеллектуальный анализ данных, прогнозная аналитика, машинное обучение, глубокое обучение, интеллектуальный анализ текста [4].

Алгоритмы интеллектуального анализа данных, такие как нейронные сети, можно научить распознавать нормальную активность пользователя и определять подозрительное поведение. Интеллектуальный анализ данных относят к диагностической аналитике. Такой метод подходит для обнаружения аномалий и обеспечения информационной безопасности, но работает только с неструктурированными данными.

Интеллектуальный анализ текста также является популярным методом для работы с данными социальных сетей. Он заключается в изучении больших коллекций документов для обнаружения новой информации или помощи в ответах на конкретные исследовательские вопросы. К его преимуществу можно отнести то, что он работает как со структурированными, так и с неструктурированными данными, а также то, что существует множество техник и алгоритмов для работы с текстом.

Предиктивная или прогнозная аналитика наиболее актуальна в социальных сетях. Процесс прогнозирования аналитики основывается на анализе исторических и текущих данных с целью создания прогнозов на будущее с помощью передовых математических и статистических методов и машинного обучения. Такая аналитика отлично подходит для оценки финансовой эффективности компании. Полученные результаты подскажут о возможных появляющихся трендах и помогут провести планирование [5]. К технологиям прогнозирования относят: деревья решений, анализ регрессии, анализ временных рядов и нейронные сети глубокого обучения.

Заключение.

Big Data, термин, который появился относительно недавно, но уже сейчас имеет огромную значимость. Он тесно связан с социальными сетями и затрагивает различные аспекты. Также

нельзя забывать о важности Big Data в информационной безопасности. А для анализа огромного количества данных будут разрабатываться различные методы и технологии, так как с ростом клиентской базы скорость поступления данных будет только расти.

Список литературы

[1] Md. Saifur Rahman. A Systematic Review Towards Big Data Analytics in Social Media / Md. Saifur Rahman, Hassan Reza // Big Data Mining and Analytics / ed. Y. Pan. – Shenzhen, 2022. – P. 228–244.

[2] Зязюлькин, С.П. Использование DQN для обучения агентов игр (Atari 2600) / С.П. Зязюлькин, С.Н. Нестеренков // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сб. материалов VI Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 20-21 мая 2020 года): в 3 ч. Ч. 2 / редкол. : В. А. Богуш [и др.]. – Минск : Бестпринт, 2020. – С. 274–280.

[3] Беляк, А. А. Анализ производительности технологии Hadoop / А. А. Беляк, С. Н. Нестеренков // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня: сб. научных статей VII Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 19-20 мая 2021 года): / редкол. : В. А. Богуш [и др.]. – Минск : Бестпринт, 2021. – С. 343–346.

[4] Jayakanna, H. S. A Study on Deep Learning / H. S. Jayakanna, Mrs. M. Raju // International Journal for Research in Applied Science and Engineering Technology – 2022 – Vol. 10, № 11 – P. 961–964.

[5] Kelleher, J. D. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies / J. D. Kelleher, B. Mac Namee, A. D'Arcy. – 2nd ed. – Boston : The MIT Press, 2020. – 856 p.

APPLICATION OF BIG DATA FOR SOCIAL NETWORKS ANALYSIS

Y.I. Bayak

Student of the specialty "Informatics and Programming Technologies" of the Faculty of Computer Systems and Networks of BSUIR

S.N. Nesterenkov

Dean of the Faculty of Computer Systems and Networks of BSUIR, PhD of Technical Sciences, Associate Professor

D.A. Zhalejko

Software Engineer of BSUIR Department of Information Technology

Department of Computer Science

Faculty of Computer Systems and Networks

Belarusian State University of computer science and Radio Electronics, Republic of Belarus

E-mail: ibayak@gmail.com

Abstract. Big Data or Big Data is a modern term that describes huge volumes of both structured and unstructured information. Social networks, as a source of Big Data, have distinctive features in their collection and analysis methods. And the results obtained are applied in various areas, ranging from information security to improving financial performance.

Keywords: Predictive analytics, information security, data mining, social network.