

УДК 004.041

Атырова Рахат Сулаймановна,
доцент, Ошский государственный университет

Атырова Рахат Сулайманова,
доцент, Ош мамлекеттик университети

Atyrova Rakhat Sulaimanovna,
docent, Osh state university

Жуманова Айгерим Эсеналиевна,
магистрант, Ошский государственный университет

Жуманова Айгерим Эсеналиевна,
магистрант, Ош мамлекеттик университети

Jumanova Aigerim Esenalievna ,
undergraduate, Osh State University

ИНФОРМАЦИОННО–АНАЛИТИЧЕСКИЕ СИСТЕМЫ ДЛЯ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ (BIG DATA)

Аннотация. Эта статья будет в первую очередь полезна для тех, кто интересуется большими данными и занимается их сбором, анализом и использованием. С применением компьютеров в трудовой деятельности большие данные будут только расти, и появится необходимость эффективной обработки, хранения и анализа этих данных. С развитием информационных технологий, появилась надобность в использовании этих данных в прогнозировании и планировании. Так как аналитика данных может дать более точный анализ, и, соответственно, более точный результат анализа может способствовать принятию правильного решения. Значимость больших данных начинает проявляться только лишь на этапе их анализа, правильное использование полезных аналитических данных повышают эффективность управления организации. А лучшее решение, в свою очередь, может означать повышение операционной эффективности, сокращение затрат и снижение рисков неправильных управленческих решений. Ведущие организации применяют аналитику данных для того, чтобы осознать риски и возможности компании с помощью больших данных, которые можно использовать в прогнозировании.

Ключевые слова: анализ данных, большие данные (Bigdata), информационно-аналитические технологии, прогнозирование, принятие управленческих решений.

КӨЛӨМДҮҮ МААЛЫМАТТАРДЫ (BIG DATA) ИШТЕП ЧЫГУУДА МААЛЫМАТТЫК-АНАЛИТИКАЛЫК СИСТЕМАЛАРДЫ КОЛДОНУУ

Аннотация. Бул макала көлөмдүү маалыматтарга кызыккандар үчүн жана биринчи кезекте аны чогултуу, талдоо жана колдонуу ишмердүүлүгү үчүн пайдалуу болот. Компьютерлер ар бир жумуш ордунда колдонулгандыктан, көлөмдүү маалыматтар өсөт жана бул маалыматтарды эффективдүү иштетүү, сактоо жана талдоо зарылчылыгы келип чыгат. Маалыматтык технологиялардын өнүгүшү менен бул маалыматты болжолдоодо жана план-

даштырууда колдонуу зарылчылыгы келип чыкты. Анткени маалыматтарды анализдөө так талдоону бере алат, тиешелүү анализдин так жыйынтыгы туура чечим чыгарууга жардам берет. Көлөмдүү маалыматтардын мааниси аларды анализдөө стадиясында гана пайда боло баштайт, пайдалуу аналитикалык маалыматтарды туура пайдалануу ишканаланы башкаруунун натыйжалуулугун жогорулатат. Туура чечим, өз кезегинде, операциялык натыйжалуулукту жогорулатуу, чыгымдарды азайтуу жана туура эмес башкаруу тобокелдигин азайтат. Алдыңкы ишканалар прогноздоодо колдонула турган көлөмдүү маалыматтар аркылуу компаниянын тобокелдиктерин жана мүмкүнчүлүктөрүн түшүнүү үчүн маалымат аналитикасын колдонушат.

Негизги сөздөр: маалыматтарды талдоо, көлөмдүү маалыматтар (Bigdata), маалыматтык жана аналитикалык технологиялар, болжолдоо, башкаруу чечимдерин кабыл алуу.

INFORMATION AND ANALYTICAL SYSTEMS FOR PROCESSING

BIG DATA (BIG DATA)

Abstract. This article is for those who are interested in big data and will be primarily useful for those who collect, analyze and use it. Big data will only grow with the use of computers in work activities, and there will be a need for efficient processing, storage and analysis of this data. There is a need to use these data in forecasting and planning with the development of information technology. Since data analytics can give a more accurate analysis, and, accordingly, a more accurate result of the analysis can help make the right decision. The significance of big data begins to appear only at the stage of their analysis; the correct use of useful analytical data increases the efficiency of the organization's management. And a better solution, in turn, can mean increased operational efficiency, reduced costs, and reduced risk of poor management decisions. Leading organizations are using data analytics to understand company risks and opportunities through big data that can be used in forecasting.

Keywords: Data analysis, big data (Big data), information and analytical technologies, forecasting, management decision making.

Введение. Постановка задачи.

В настоящее время невозможно представить какую-либо организацию, которая не использует информационные технологии в своей деятельности. С применением компьютеров в повседневной жизни и в трудовой деятельности накопилось очень много информации, некоторые из них мы, отструктурировав, сгруппировали и объединили в базу данных, которую затем храним долгими годами. Эти данные настолько велики, что ни один из традиционных инструментов управления данными не может их эффективно хранить или обрабатывать. Важен не только объем данных, но и то, как организации используют данные. По некоторым источникам, люди производят 2 квинтилли-

она данных каждый день. Например, одна только Нью-Йоркская фондовая биржа создает около одного терабайта новых торговых данных в день. Платформы социальных сетей также вносят большой вклад в сбор данных. Кроме того, авиакомпании также генерируют много петабайт данных. [1]

Многие организации обеспокоены тем, что количество накопленных данных становится настолько большим, что трудно найти наиболее ценную и подходящую информацию. Как эффективно мы можем использовать накопленные данные? С развитием информационных технологий, появилась необходимость в использовании этих данных в прогнозировании и планировании. Со временем оказалось, что аналитика данных яв-

ляется серьёзным и нужным направлением, что возник отдельный вид информационных систем – информационно-аналитические системы.

Цифровизация открывает новые разделы информационных технологий, такие как, облачные технологии, большие данные, аналитика данных и другие.

Данные накапливались годами и их стало очень много, и появилась задача использования этих сохраненных данных т.е. обрабатывать эти данные таким образом, чтобы они могли нам «добывать» полезную информацию.

Чтобы проводить аналитику над данными, они нужны в больших количествах, т.е. данные должны накапливаться годами и их должно быть очень много, чтобы эффективно сделать аналитику и извлечь из этих сырых данных только полезные. В литературе они известны как большие данные или BigData. Это направление включает в себя много разделов, такие как DataMining, OLAP, DataWarehouse, бизнес аналитика (BI) и т.д.

Современные предприятия просто не могут быть конкурентоспособными без того или иного использования больших данных для поддержки операций, управления, планирования или просто принятия основных решений. Большие данные могут быть проанализированы для получения информации, которая поможет принять более эффективные решения и принять стратегические решения в управлении. А также на всех уровнях власти большие данные используются для вовлечения граждан и для руководства выработкой политики, преследующей интересы общественности и общества в целом.

Большие данные могут стать для бизнеса и общества такими же полезными, как и Интернет. Они могут дать более точный анализ, и, соответственно, более точный результат анализа может способствовать принятию правильного решения. А лучшее

решение, в свою очередь, может означать повышение операционной эффективности, сокращение затрат и снижение рисков.

Аналитика больших данных и прогнозная (предиктная) аналитика могут помочь в развитии организации с помощью наборов потребительских данных, которые анализируются, группируются и очищаются от всей ненужной информации. Прогнозный анализ больших данных позволяет компаниям заглянуть в будущее и расти, основываясь на прошлом и настоящем опыте и заранее спланированных будущих направлениях.

По мнению разработчиков, прогнозная аналитика больших данных относится к расширенной аналитике. Он способен прогнозировать будущие результаты с помощью исторических данных, статистического моделирования, интеллектуального анализа данных и машинного обучения. Предприятия используют прогнозную аналитику, чтобы понять свои риски и возможности с помощью шаблонов данных, которые можно предсказать. [2, 6]

Чтобы предсказать будущие события, прогнозирующая аналитика с помощью статистического анализа выявляет значимые модели больших данных. Аналитика данных может применяться к неизвестным данным в настоящем, прошлом и будущем. С помощью анализа больших данных, прогнозная аналитика способна предоставить ценную бизнес-аналитику для любой организации, вне зависимости от формы управления, так как любая деятельность хранит большие данные в своей отрасли. И аналитика данных, основываясь на накопленных структурируемых данных, проводит вычисления и извлекает полезные данные, которые используются в прогнозировании. Как правило, организации могут применять прогнозную аналитику трех типов:

- прогнозное моделирование;
- описательное моделирование;
- моделирование принятия решений.

По сравнению с другими направлениями и информационными технологиями, аналитика данных появилась недавно, но уже можно с уверенностью сказать, что в будущем это направление будет только развиваться. Перспективы аналитики данных очень велики, с переходом на цифровую форму управления, документооборота, отчетности и цифровой трансформации других процессов. С применением компьютеров во всех сферах деятельности большие данные будут только расти, и вместе с ними будут разрабатываться новые технологии для более эффективного сбора, хранения и анализа данных, поскольку мир трансформации, основанной на данных, движется вперед со еще большей скоростью. Оцифровка, структурирование данных и, в дальнейшем, хранение их в OLAP кубах будет эффективным способом для дальнейшей обработки больших данных. Чтобы правильно структурировать, а затем их анализировать, нужны будут специалисты. Уже сейчас чувствуется нехватка профессиональных аналитиков в области больших данных.

В нашей стране для эффективного управления государством было бы наилучшим выходом использование аналитических центров, которые работали бы в полную мощь с применением современных программ аналитики в области оптимального стратегического принятия решения не только в администрации президента, но и от кабинетов министерств вплоть до МСУ (местного самоуправления). Собственно говоря, с переходом в цифровую трансформацию Электронного правительства Кыргызской Республики, использование анализа накопленных данных (в ЦОД-центр обработки данных) в принятии управленческих решений на разных уровнях управления государством является востребованным и необходимым инструментом в настоящее время.

Термин «большие данные» в большинстве источников связывают с именем Клиффорда Линча, который в 2008 году в журнале Nature опубликовал ста-

тью на тему «Как могут повлиять на будущее науки технологии, открывающие возможности работы с большими объемами данных?» А в 2010 году стали разрабатываться первые программные продукты по обработке больших данных. С 2011 года компании IBM, Oracle, Microsoft, Hewlett-Packard, EMC и Gartner заинтересовались большими данными и стали вносить свой вклад в развитие аналитики данных. С развитием облачных технологий анализ больших данных стал возможным в использовании не только для крупных организаций, но стал реально доступным средством для малого и среднего бизнеса.

Что касается ответа на вопрос: Что такое аналитика больших данных? Компания Microsoft ответила, что аналитикой больших данных называются методы, инструменты и приложения, используемые для сбора, обработки и получения аналитических сведений из разнородных, объемных наборов данных, накапливающихся с высокой скоростью. [3]

Я постаралась найти более точное определение, полагаясь на другие источники: можно сказать, что аналитика больших данных — это сложный процесс изучения больших данных для выявления информации, который выявляет скрытые закономерности, корреляции, рыночные тенденции и предпочтения клиентов, которые могут помочь организациям принимать обоснованные управленческие решения. [4]

В широком смысле технологии и методы анализа данных дают организациям возможность анализировать наборы данных и выявлять новую полезную информацию, которая нужна для принятия решений. Важность аналитики больших данных в том, что организации могут использовать аналитические системы и программное обеспечение анализа больших данных для принятия управленческих решений. На основе накопленных данных, которые могут улучшить результаты прогнозирования и помочь принятию оптимального и эффективного реше-

ния без рисков. Преимущества могут включать в себя более эффективный маркетинг, новые возможности получения дохода, персонализацию клиентов и повышение операционной эффективности. [7]

Аналитики данных работают по следующей схеме (как утверждает BitnewsToday) : специалисты по анализу данных, специалисты по прогнозному моделированию, статисты и другие специалисты в области аналитики данных собирают, обрабатывают, очищают и анализируют растущие объемы структурированных данных, например, о транзакциях, а также другие формы данных, которые не используются обычными программами бизнес-аналитики.

Аппаратное и программное обеспечение для прогнозной аналитики, которое обрабатывает большие объемы сложных данных использует машинное обучение и статистические алгоритмы для прогнозирования результатов будущих событий. Организации успешно используют инструменты прогнозирующей аналитики для обнаружения мошенничества, маркетинга, оценки рисков и операций. Например, для прогнозирования спроса использовались различные методы статистического анализа, включая анализ временных рядов и регрессионный анализ [5]. Благодаря достижениям в области информационных технологий и повышению эффективности вычислений, аналитика больших данных стала средством получения более точных прогнозов, которые лучше отражают потребности клиентов, облегчают оценку производительности, повышают эффективность, сокращают время для принятия решений и автоматизацию процессов [6].

Сохранить большие данные с подручными средствами, например, как программными продуктами Microsoft (Excel, Access, Power Point или даже с помощью СУБД SQL) невозможно реализовать. Для эффективной обработки больших данных нужны специальные средства как Hadoop & MapReduce,

NoSQL базы данных, Инструменты класса DataDiscovery и т.д. Как и все новшества аналитика данных в основном применяются в коммерческих организациях, так как большинство программ аналитики данных стоит не малых денег. Но в тоже время есть инструменты условно бесплатные или программы с открытым исходным кодом, например, Knime, Open Refine, R-Programming, Rapid Miner и др. [8]

Выводы:

1. Аналитика данных является оптимальным решением, которое успешно можно применять в любой деятельности, вне зависимости от формы управления: будь то государственная организация или бизнес сфера.

2. Чтобы эффективно проводить аналитику больших данных, нужно уметь использовать инструменты для структурирования данных. Оптимально хранить данные с помощью OLAP и Data Warehouse. Также данные должны быть очищены от дублирования, неточных входных данных, системных ошибок и других видов отклонений, чтобы впоследствии получить более точные результаты аналитики больших данных.

3. Необходимо, по моему мнению, внедрять отчеты аналитических центров в разные уровни управления государством и бизнеса, так как Кыргызская Республика планирует реализовать полный переход цифровой трансформации на государственном уровне в 2025 году. Нужно сейчас думать, как сохранить данные в оптимальной форме, чтобы легко было их в дальнейшем обрабатывать и использовать в аналитических центрах.

4. Способствовать подготовке и вовлечению специалистов в области анализа данных, в которую входит специалисты интеллектуального анализа данных и текста, искусственного интеллекта, машинного обучения, предиктивной аналитики и глубокого обучения.

Литература

1. Complete Guide to Predictive Analytics and Big Data Analytics. Nazar Kwartalny [Электронный ресурс]. – Режим доступа: <https://inoxoft.com/blog/complete-guide-to-predictive-analytics-and-big-data-analytics/> Опубликовано: 25.08.2021. (дата обращения: 20.04.2022).
2. *Майер-Шенбергер В., Кукьер К.* Большие данные: Революция, которая изменит то, как мы живем, работаем и мыслим. – Манн, Иванов и Фербер, 2014 -240с.
3. Что такое аналитика больших данных? [Электронный ресурс]. – Режим доступа: <https://azure.microsoft.com/ru-ru/overview/what-is-big-data-analytics/#importance-of-data-analytics> (дата обращения: 30.01.2022).
4. *Бабанов, А. Б.* Перспективы внедрения больших данных в бизнесе / А. Б. Бабанов, В. В. Кадацкая. — Текст : непосредственный // Молодой ученый. — 2021. — № 28 (370). — С. 174-176. — URL: <https://moluch.ru/archive/370/83188/> (дата обращения: 20.05.2022).
5. Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. [Электронный ресурс]. – Режим доступа: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00329-2> (дата обращения: 05.05.2022).
6. Wang G, Gunasekaran A, Ngai EWT, Papadopoulos T. Big data analytics in logistics and supply chain management: certain investigations for research and applications. Int J ProdEcon. 2016;176:98–110. <https://doi.org/10.1016/J.IJPE.2016.03.014>.
7. Herodotou H. et al. Starfish: A Self-tuning System for Big Data Analytics //CIDR. – 2011. – Т. 11. – p.261-272.
8. Осторожно, данные собираются: как технологии bigdata меняют мир [Электронный ресурс]. – Режим доступа: <https://finparty.ru/opinions/166219/>. (дата обращения: 21.12.2021).
9. *Кушнир, Е. А.* Противостояние XXI века / Е. А. Кушнир, Л. А. Телегина. в 2-х томах. Т. 1. – Саров: ФГУП «РФЯЦ-ВНИИЭФ», 2019. – 354 с