

ОСНОВНИ ТЕХНОЛОГИИ И ПРИЛОЖНИ НАСОКИ ЗА ДИГИТАЛНА ТРАНСФОРМАЦИЯ НА БИЗНЕСА В СРЕДА НА ГОЛЕМИ ДАННИ

**Ас. д-р Станимира Йорданова¹,
Проф. д-р Камелия Стефанова²**

Резюме: В съвременното развитие на бизнеса дигиталната трансформация се налага като предизвикателство за управлението на големи данни. В среда на постоянно нарастващ обем от данни основните критични фактори за развитие на бизнеса са свързани с обработване и анализ на големи данни. Технологиите за обработка на данните с отворен код използват иновативен подход при разработването им в партньорство между разработчиците, който осигурява прозрачност, достъпност и постоянно усъвършенстване.

Основната цел на статията е да разгледа популярни технологии за обработка на големи данни с отворен код и текущи насоки за бизнес приложения в средата на големи данни.

Ключови думи: големи данни, дигитална трансформация, технологии за обработка на големи данни.

JEL: O33, C88.

¹ Катедра "Информационни технологии и комуникации", Университет за национално и световно стопанство, syordanova@unwe.bg

² Катедра "Информационни технологии и комуникации", Университет за национално и световно стопанство, kstefanova@unwe.bg

Въведение

През 50-те години на миналия век цифровизацията (digitization) започва с конвертиране на аналогови сигнали или информация от всякаква форма в цифров формат, който може да бъде разбран от компютърни системи или електронни устройства и вземане на аналогов процес и превръщането му в цифрова форма без каквито и да било промени в процеса (Gartner Glossary). Последващото развитие на интернет технологиите през 90-те години на миналия век и електронната търговия през 2000 година промени начина, по който хората работят, пазаруват, пътуват, обучават, управляват и изобщо живеят. Широкото разпространение и използване на интелигентни устройства като смартфони, таблети и мобилни приложения създаде нова пазарна реалност, към която търговците трябва да се адаптират, за да останат конкурентоспособни. Дигиталните технологии позволяват, клиентите винаги да са в интернет и да очакват съдържание, съобразено с това, което правят по всяко време, навсякъде и на устройство по техен избор. „Нарастващото навлизане на дигиталните технологии в обществото със свързаните с това промени в поведението на хората” се определя от Gimpel, H. и Röglinger, M. (2015) като дигитализация. Но тяхното определение за дигитализация отразява само социалния аспект на явлениято. Новото поведение на клиентите принуждава бизнеса да въвежда нови технологии с цел достигане на по-високо ниво на развитие на бизнес процесите, продуктите и услугите.

Процесът на адаптация на бизнеса към новата пазарна реалност в следствие на променящото се поведение на клиентите отразява бизнес аспекта на дигитализацията. Gartner разглежда дигитализацията от бизнес гледна точка: „използването на дигитални технологии за промяна на бизнес модели и предоставяне на нови възможности за реализиране на приходи“ и допълва, че това е „процесът на преминаване към дигитален бизнес“.

Kurt Brand (2017) надгражда понятието дигитализация, като го определя като: „иновация на бизнес модела и процесите, които водят до дигитални (бизнес) възможности, например чрез дигитални платформи, осигуряващи екосистеми за създаване на стойност, основани

ОСНОВНИ ТЕХНОЛОГИИ И ПРИЛОЖНИ НАСОКИ ЗА ...

на услуги, базирани на данни“. Дигитализацията според Brand обхваща промяната на **бизнес процесите**, предизвикана от новите изисквания на клиентите чрез използване на дигитални технологии; използване на **нови източници на данни**, които да се анализират непрекъснато чрез интегриране на сензори в продуктите, процесите и системите; **промяна на организационната култура** и ориентирането ѝ към клиентите и създаване на **дигитална екосистема**, извън организацията, обхващаща клиенти, доставчици, партньори дори и конкуренти.

Процесът на заместване на традиционните съществуващи бизнес модели с иновативни бизнес модели, използвайки напреднали технологии, който предизвиква фундаментална промяна в икономиката, организацията и обществото се нарича **дигитална трансформация**. Пример за иновативен бизнес модел в областта на търговията, който променя достъпа на клиенти и доставчици до пазара, е дигиталната екосистема на Амазон. Бизнес моделът на Амазон и постигнатата пазарна сила принуждава търговците да използват платформата, за да увеличат значително продажбите. Тъй като Amazon разполага и контролира всички данни на търговците, може да поеме напълно бизнеса им, ако това се вписва в стратегията на компанията. Освен това голямото разнообразие от транзакции предоставя на Амазон огромно количество данни, които те могат да използват за прогнозиране, контролиране или дори манипулиране на поведението на клиентите и доставчиците в интерес на Амазон.

Бизнес фактори, които се превръщат в двигатели на трансформацията в бизнеса, са:

- Предоставяне на последователна и персонализирана продуктова гама на клиентите по всички възможни канали;
- Развиване на конкурентно предимство чрез създаване на уникални възможности за подобряване на удовлетвореността на клиента;
- Получаване на задълбочено разбиране за поведението на клиентите с цел повишаване на лоялността;
- Провеждане на ефективни и целенасочени маркетингови кампании, резултат от анализи на данни от множество източници;

- Намаляване на разходите чрез оптимизиране на инвентаризацията и управлението на веригата за доставки.

Интернет на нещата (Internet of Things), изкуствен интелект (Artificial Intelligence), виртуална реалност (Virtual reality), облачни технологии, социални медии са основни **технологични фактори**, които движат процесите на дигитализация в бизнеса и предоставят възможности за достъп до нови **данни с голям обем**, които да се анализират и използват в процеса на вземане на управленски решения.

Днес моделите на дигитална трансформация се изграждат около **големите данни**, технологиите на интернет на нещата и управлението и анализа на големи данни. Според проучване на Dresner Advisory Group на пазара на Big Data технологии през 2018 анализът на големи данни е посочен като критичен фактор за дигиталната трансформация от телекомуникационната, рекламната и застрахователната индустрия и като съществен фактор в развиването на здравеопазването и търговията. 80% от всички предприятия посочват, че големите данни имат най-важна роля в техните стратегии за развитие на бизнес интелигентност (Columbus, L., Forbes 2018).

Големите данни стават ключов двигател за дигитална трансформация, защото откриват възможности за достъп до нови клиенти, канали на разпространение на продукти и пазари, както и за подобряване на управлението на бизнеса. В средата на постоянно нарастващ обем от данни, основно предизвикателство пред бизнеса са събирането, съхраняването, обработването, интеграцията и анализа на големи данни.

1. Големи данни в бизнеса

Големи данни са данни с голям обем, висока скорост и/или разнообразен формат, които изискват икономически ефективни и иновативни форми на обработка и позволяват по-добро взимане на решения и автоматизация на процесите. През 2001г. Gartner дефинира трите характеристики на големите данни: “volume”, “velocity” и “variety” (обем, скорост и разнообразие на данните). През годините характеристиките

ОСНОВНИ ТЕХНОЛОГИИ И ПРИЛОЖНИ НАСОКИ ЗА ...

на големите данни се разрастват, като през 2017 г. Том Шафър (Data Scientist) от компанията „Elder Research“ представя актуализиран списък с 42 V's на големите данни.

- Обемът (Volume) е най-известната характеристика на „големите данни“. Данните нарастват с 40% годишно, като се очаква да достигнат близо 40 ZB през 2020г. (AT Kearney). Такива големи обеми от данни могат да се обработват, управляват и анализират само от специализирани технологии за големи данни.

- Скоростта (Velocity) се отнася до скоростта, с която се генерират, произвеждат, създават или обновяват и обработват данните с цел извличане на полезно знание.

- Разнообразието (Variety) обхваща различните типове (структурирани, полуструктурирани и неструктурирани данни) и източници на данните.

Останалите характеристики на големите данни (Variability – променливост, Veracity – вярност, Validity – валидност и т.н) допълват основните предизвикателства и предимства в прилагането на големи данни.

В литературата данните се категоризират от гледна точка на техния формат (структурирани и неструктурирани), обем и скорост, с която се произвеждат и използват (Rozados и Tjahjono, 2014; Deloitte, 2012, Hortonworks/Terradata).

Структурирани вътрешни данни, генерирани от бизнес процесите на компанията от Системи за планиране на ресурсите (ERP), Системи за управление на връзките с клиентите (CRM) и др.:

- Данни за продажби от клиенти (POS и исторически данни);
- Данни за/от доставчици;
- Данни за складови и транспортни дейности;
- Данни от фактури и трансакции;
- Данни от клиентски профили, договори и др.;
- Маркетингови данни (цени, ценообразуване и др.);
- Данни за управление на доставките;
- Данни от електронни магазини;
- Финансови и бюджетни данни;
- Данни за мърчандайзинг;

- Уеблогове, трафични данни от клиенти, потребителски кликове от страници на компанията и др.

Структурирани външни данни, генерирани от външни източници за организацията:

- Сензорни данни от IoT устройства;
- GPS данни и данни от смарт телефони;
- Данни от игри (Gamification data);
- Данни от симулации и други средства за виртуално и интерактивно взаимодействие с продуктите;
- Данни от средства за виртуална реалност (virtual reality), разширена реалност (augmented reality) и смесена реалност (mixed reality);
- Агрегирани данни (Syndicated) от специализирани в областта на търговията източници като например бизнес асоциации, които агрегират данни от много търговци и публикуват структурирани данни в областта;
- Данни от дистрибутори и доставчици;
- Данни за времето;
- Макро- и микроикономически данни от публични източници (демографски, статистически и др.).

Неструктурирани вътрешни данни, генерирани от електронна поща, уеб, колцентрове и други вътрешни бизнес процеси, основно в текстови, аудио, видеоформат или изображения:

- Данни от пазарни проучвания (чрез въпросници и др. инструменти);
- Данни от центрове за обслужване на клиентите и колцентрове;
- Данни от страници на компанията в социални медии;
- Данни от интервюта с кандидати за работа;
- Данни от оценки за работата на служителите, както и документи, съдържащи обратна връзка от служителите;
- Данни от документи (отчети, дневници, проучвания, електронна поща);

ОСНОВНИ ТЕХНОЛОГИИ И ПРИЛОЖНИ НАСОКИ ЗА ...

• Данни от социални бизнес мрежи на доставчици, дистрибутори и търговци, от мрежи на професионални работодателски асоциации и мрежи за отворени иновации и сътрудничество.

Неструктурирани външни данни, генерирани от източници извън организацията:

- Данни от приложения, използвани от клиенти;
- Доклади за качество и безопасност на продуктите;
- Професионални продуктови ревюта от експерти;
- Данни от новинарски медии;
- Коментари, блогове, ревю сайтове, в които потребителите дават своето мнение от използването на продукти и услуги (user-generated content);
- Видео- и социални мрежи и платформи;
- Изображения и разпознаване на изображения;
- Видеопроследяване и мониторинг от системи за наблюдение в магазини.

Структурираните данни, идващи от вътрешни и външни за организацията източници, са с малък обем и се съхраняват в реляционни бази данни и/или складове за данни. Неструктурираните данни са без модел във формат на текст, видео, изображения, звук. Те се съхраняват в NoSQL бази данни, Data lakes или други хранилища. Тъй като основните източниците на тези данни са извън организацията, те са с най-бързо нарастващ обем.

Големите данни, характеризиращи се с постоянно нарастващ обем и разнообразие от формати, изискват специални технологии за управление на данните.

2. Технологии и инструменти за управление на големи данни

Утвърдени етапи на обработка на големи данни са извличане и събиране на данни, съхраняване и обработка на данни, подаване на заявки към данните, анализ и визуализация и са представени на Фигура 1 (Gill, 2017).



Фигура 1. Етапи за обработка на големи данни

Технологиите, които реализират изпълнението на процесите във всеки етап, са многобройни както комерсиални, така и с отворен код.

Традиционните комерсиални решения за интеграция на данни са солидни, надеждни и подкрепени с екипи за поддръжка и развитие. Въпреки това тези продукти обикновено са изградени на основата на традиционна архитектура, която може да не се адаптира лесно към голямата среда за данни и внедряването им да изисква значителни разходи.

За разлика от тях технологиите с отворен код предлагат гъвкавост на лицензи и спестяване на разходи. Но най-голямото им предимство е иновативният партньорски подход на разработване и развитие

на технологиите и пример за това са Apache проектите. Тези предимства на технологиите с отворен код водят до тяхната популярност и често използване. В резултатите от изследване на Dresner Advisory Group от 2017 г. на пазара на технологии за големи данни, Spark SQL, Hive, HDFS, Amazon S3, HBase, Impala, MongoDB, Cassandra се посочват като предпочитани технологии за достъп и обработка на големи данни. В статията са представени популярни технологии с отворен код, които често се използват за реализиране на етапите от методологията за обработката на големи данни.

Извличане и събиране на данни

Процесът на извличане на данни започва с приоритизиране на източниците на данни, валидиране и транспортиране на данните до правилното местоназначение в пакетен режим или в реално време. В допълнение към извличането, събирането и интегрирането инструментите помагат, данните да се променят и формират за целите на съхранение и анализ. Основни фактори, които имат значение в процеса на събиране на данните са скорост на данните (скорост, с която данните постъпват от различни източници); размер на данните (огромен и увеличаващ се обем данни); честота на данните (в реално време или на партиди/интервали) и формат на данните (структурирани, полуструктурирани, неструктурирани). За целта може да се ползват различни инструменти за извличане на големи данни с отворен код, някои от които са:

- Apache Kafka е система за обработка на големи данни в реално време. Kafka обработва входящи потоци от данни независимо от техния източник и дестинация. Области на приложение са проследяване на дейностите на уебсайт; обработка на потоци; събиране и мониторинг на показатели.

- Apache Flume е система за събиране, агрегиране и прехвърляне на потоци от данни от различни източници (логове, сензори, машини, социални медии) към централизирано хранилище на данни, като например HDFS и Hbase. За разлика от Kafka, който може да обработва и наблюдава данни в разпределени системи, Flume събира данни от разпределени системи, за да ги прехвърли в централизирано хранилище на данни. Когато са конфигурирани правилно, Apache Kafka и Flume са много надеждни с гаранции за нулева загуба на данни.

- Apache Nifi е инструмент, който се използва за зареждане на данни от различни източници и прехвърлянето им към друг източник.
- Apache Sqoop е инструмент за двупосочен трансфер на данни между Hadoop и SQL бази данни.
- Cloudera Morphlines е система с отворен код за интегриране, изграждане и промяна на приложенията за обработка на Hadoop, които извличат, трансформират и зареждат данни в HBase или HDFS или корпоративни хранилища за данни (data warehouses). Използва се за приложения в реално време, както и за приложения в пакетен режим.

Съхраняване на данни

Този етап обхваща съхраняването на структурирани данни в системи за управление на релационни бази данни и неструктурирани данни в специализирани файлови системи като Hadoop Distributed File System (HDFS) или NoSQL база данни.

- Hadoop Distributed File System (HDFS) е класическата файлова система за големи данни. Hadoop осигурява паралелната обработка на данни между изчислителните възли с цел увеличаване на скоростта на изчисленията и намаляване на латентността. Данните се записват веднъж и след това се четат многократно за разлика от файловите системите, при които данните постоянно се четат и записват. HDFS е устойчива среда срещу повреди в сървърите, защото данните във вид на блокове се репликират навсякъде в клъстера. Тази технология е популярна благодарение на здравината и неограничения мащаб на хардуера. Въпреки това изисква специализиран набор от умения и сложна интеграция на безброй компоненти с отворен код.
- HBase е нерелационна система за управление на бази данни във формат колони, която е изключително ефективна и бърза при четене и писане на големи набори от данни в реално време. Работи с HDFS и поддържа MapReduce заявки.
- MongoDB – е нерелационна база данни, която съхранява всякакъв тип данни в отделен документ. Достъпът до документи е чрез индексирани и осигурява бърз отговор на заявки. Скоростта на MongoDB е 100 пъти по-бърза от релационната база данни. Някои обаче критикуват използването му като първокласна система за съхранение на данни

поради ограничените си аналитични възможности и липсата на поддръжка на транзакционни данни.

- Amazon Simple Storage Service (Amazon S3, AWS) е услуга за съхранение на данни като обекти. Всеки обект се съхранява като файл с метаданни и идентификационен номер. Приложенията използват този идентификационен номер за достъп до обекта. Услугата предлага сигурност, производителност, евтино и гъвкаво съхранение и бърза реализация. Някои автори посочват като недостатъци скорост на извличане на данните и трудна настройка за комуникация между софтуера за управление на данни и AWS.

Обработка и управление на данните

Процесите на обработка на данни трансформират, категоризират и оптимизират данните (пакетно или в реално време) с цел улесняване на последващия анализ.

- Apache Sqoop е система за пакетна обработка на данните, като прехвърля данни между Apache Hadoop и структурирани хранилища на данни. Sqoop може да се използва за импортиране на данни от външни структурирани хранилища на данни в HDFS или свързани системи като Hive и HBase. Sqoop може да се използва за извличане на данни от Hadoop и експортиране към външни структурирани бази данни.

- Apache Spark е система за обработка на потоци от данни в реално време и е разработен, за да се справи с недостатъците на Hadoop. Например той може да обработва както пакетни данни, така и данни в реално време и работи 100 пъти по-бързо от MapReduce. Spark осигурява възможности за обработка на данни в паметта и работи с HDFS и Apache Cassandra.

- Apache Flink е рамка с отворен код за обработване на разпределен поток. Някои от неговите функции са: стабилен и устойчив на откази и може безпроблемно да се възстанови от повреди; работи на хиляди възли с отлични пропускателни и латентни характеристики; API и библиотеки за пакетно, стрийминг обработване, машинно обучение и графична обработка. Може да се приложи за оптимизиране на резултатите от търсенето в електронна търговия в реално време; мрежово/сензорно наблюдение и откриване на грешки; ETL при БИС.

- Apache Cassandra е една от технологиите, които стоят зад огромния успех на Facebook, тъй като позволява да се обработват структурирани масиви от данни, разпределени в огромен брой възли по целия свят. Съхранява данните в колони, но за разлика от традиционните бази данни, където имената на колони се състоят само от метаданни, в имената на колоните при Cassandra се съдържат действителните данни. Предимства на технологията са голямата линейна мащабируемост, постоянна репликация по възлите; лесно добавяне и премахване на възли от работещ клъстер; висока толерантност към грешки.

- Apache SAMOA (Scalable Advanced Massive Online Analysis) е платформа с отворен код за обработка на големи потоци от данни. Платформата осигурява набор от разпределени стрийминг алгоритми за най-често срещаните задачи за извличане на знания от данни и машинно обучение като класификация, клъстеризация и регресия, както и разработване на нови алгоритми.

SQL заявки

- Apache Hive е софтуер за съхранение на данни, който улеснява четенето, писането и управлението на големи набори от данни, които се намират в разпределено хранилище, използвайки SQL. Hive предоставя SQL-подобен интерфейс за заявки към данните, съхранявани в различни бази данни и файлови системи, които се интегрират с Hadoop.

- Apache Impala е аналитична база данни с отворен код за Hadoop. Използва се от системи като Cloudera, MapR, Oracle и Amazon. Поддържа SQL заявки за данни, независимо дали се съхраняват в HDFS или HBase в реално време.

Инструменти за извличане на знания от данни

В етапа на прилагане на аналитични средства за извличане на знания от големите данни се използват библиотеки за машинно обучение и инструменти за извличане на знания от данни (Data Mining). Извличането на знания от данни (Data Mining) е процес, при който се откриват скрити модели в големи обеми данни с цел извличане на ново знание от данните. Този процес използва специфични алгоритми, статистически анализ, изкуствен интелект и бази данни, за да извлича ин-

ОСНОВНИ ТЕХНОЛОГИИ И ПРИЛОЖНИ НАСОКИ ЗА ...

формация от огромни масиви от данни и да ги преобразува в разбираема форма. Някои от инструментите за извличане на данни, широко използвани с големи данни от индустрията, са:

- MLib е библиотеката на Spark за машинно обучение, която предоставя алгоритми и методи, които покриват целия процес на машинно обучение.

- Apache Mahout е рамка за създаване на алгоритми за машинно обучение. Фокусира се върху клъстериране на данни и класификация. Mahout е написан на JAVA и включва библиотеки на JAVA за извършване на математически операции като линейна алгебра и статистика. Apache Mahout има следните основни характеристики: разширяема среда за програмиране, предварително направени алгоритми, среда за математическо експериментиране.

- Rapid Miner е софтуерна платформа за анализ на данни, която осигурява интегрирана среда за подготовка на данни, машинно обучение, извличане и обработка на текст и прогнозен анализ. Той е една от водещите системи за извличане на данни. Програмата е изцяло написана на Java език за програмиране и предлага възможност за експериментиране с огромен брой оператори и удобен графичен интерфейс за бързо моделиране и извличане на знания.

- KNIME (Konstanz Information Miner) е платформа за анализ на данни с отворен код, която прави възможно изграждането на модели за предсказване от потребители, които нямат опит в програмирането. Платформата включва множество модули и готови за употреба примери и масив от интегрирани инструменти и алгоритми за анализ на данните. KNIME е конкурент на RapidMiner в областта на анализа и извличането на знания от данни.

- Python е език за програмиране с общо предназначение, който може да се изпълнява в Windows и Linux среди. Python е широко използван в научните и бизнес среди, защото се състои от голям брой библиотеки за анализ, визуализацията, машинното обучение.

- Rattle GUI е отворен и свободен софтуерен пакет, осигуряващ графичен потребителски интерфейс за извличане на данни с помощта на статистически език за програмиране R. Rattle също може да се използва като средство за изучаване на R, за статистически анализ

или разработване на модели за предсказване, като предоставя методи за обучение, валидиране и тестване.

Бизнес интелигентност и визуализация

Извличането на данните събира суровите данни, необходими за анализ, интеграцията ги смесва всички заедно, а обработката ги оптимизира и организира с цел анализ и визуализация. Целта на бизнес интелигентния анализ е да се отговори на бизнес въпроси и да се подобри взимането на управленски решения. Използвайки бизнес интелигентните технологии, бизнес анализаторите могат да обобщават данните и да ги изследват в детайли, като създават табла за управление и интерактивни визуализации.

Разнообразието от БИ софтуер е голямо, някои имат безплатни версии (QlikSense, Microsoft Power BI), други са с високи цени (Tableau, QlikView), но включват безплатни допълнителни услуги. Някои от тях са по-подходящи за конкретни приложения, докато други изискват по-обширни умения за кодиране или опит в областта на анализа на данните.

Популярен БИ софтуер с отворен код е BIRT. Според Eclipse Foundation е изтеглена повече от 12 милиона пъти и е подкрепена от повече от 2,5 милиона разработчици от най-малко 157 страни. BIRT се използва от компании като Cisco и IBM и може да бъде вградена и в много други бизнес приложения, но изисква обширни умения за кодиране, така че организациите, които нямат тези ресурси, трудно биха го използвали.

Chartio е облачна БИ услуга, която предоставя лесен и директен достъп до данни в CSV или Google Sheets, както и интеграция с други облачни приложения. Looker е също облачна БИ услуга за анализ на големи данни чрез SQL.

3. Основни приложни аспекти на големи данни в бизнеса

Представените технологии предлагат възможности за събиране, обработване и анализиране на големи данни за решаване на различни бизнес проблеми в компаниите. Основните аспекти на приложе-

ние на големите данни в бизнеса са свързани с обезпечаване на оперативните дейности, управление на взаимоотношенията с клиентите, маркетинг и ценообразуване, откриване на измами.

Оперативна аналитичност (Operational Analytics)

Оперативната аналитичност е свързана с използването на големи данни в системи, които анализират машинни и сензорни данни с цел откриване на отклонения и проблеми в оперативните дейности. Данните в тези системи постъпват от сензори и програми за наблюдение на работата на оборудването и се използват за анализ и превантивна поддръжка, която може да помогне за предотвратяване на инциденти или прекъсвания. В много производствени процеси е важно да се предвижда оставащият оптимален живот на системите и компонентите, за да се гарантира, че работят в рамките на спецификациите.

Оптимизирането на производствените линии може да намали разходите и да увеличи приходите. Големите данни могат да помогнат на производителите да разберат потока от стоки в производствените линии. Анализът на данните може да разкрие причините за увеличаване на времето за производство.

Оптималното функциониране на мрежите е от съществено значение в телекомуникационния бизнес. Анализът на използване на мрежата може да помогне на компаниите да идентифицират области с излишен капацитет и да пренасочат трафика, ако е необходимо. Големите данни могат да подпомогнат планирането на инвестиции в инфраструктурата и създаването на нови услуги, които отговарят на клиентските изисквания.

Анализ на клиенти (Customer Analytics)

- *Изследване на потребителското поведение*

Бизнесът продава своите продукти по различни канали, уебсайтове и дори директно от своите доставчици. Това генерира голям обем от данни за взаимодействия между търговците и техните клиенти. Благодарение на дигиталните технологии днес клиентите са свободни за избиране и купуват продукт в магазина или онлайн, като имат възможности да сравняват цените онлайн и дори да купуват от уебсайта на конкуренти. Използването на данните от източниците в реално време

се оказва трудно, а само анализът на историческите данни за поведението на клиентите е недостатъчен. Чрез корелация на данните за продажбите с фактори като време на покупка, демографска статистика, артикули в пазарната кошница, прилагани отстъпки или купони и външни фактори като сезонност търговците могат да научат повече за поведението на клиентите. С интегрирането на данни от камери и сензори могат да разберат по-добре моделите на покупка и да подобрят предлаганите промоции.

Анализирането на всички персонални данни за клиентите, данни от взаимодействия с клиентите и данни от социални медии би довело до оптимизиране на клиентския опит по всички комуникационни канали, подобряване отношението към бранда, ангажиране на крайния потребител на лично ниво чрез персонализиране на оферти, ефективно предлагане на правилния продукт или промоция в точното време.

- *Подобряване на програмите за лоялни клиенти*

Програмите за лоялни клиенти са широко използвано средство за привличане на клиенти обратно в магазините. Предизвикателството обаче сега е да привлече клиента обратно в магазина, когато конкурентът предлага програма, която по същество е същата. Програмите за лоялност предоставят ценни данни на клиента за идентифициране на информация, която може да бъде екстраполирана към по-голяма непознатата аудитория. Освен това търговците могат да си партнират помежду си, като създадат уникални оферти, базирани на комбинации от покупки като допълнение към програмите за лоялност.

- *Изследване на потребителското мнение*

Развитието на дигиталните и уебтехнологиите предостави на клиентите много повече начини да изразяват мнението си от използването на продуктите чрез туитове, харесвания, споменавания, препоръчвания. Тъй като мненията са текст, използването на методи от анализ на текст (Text Mining) в системите за големи данни подпомага бизнеса да открива модели в неструктурирани данни и да идентифицира фенове, критици, потребители, които влияят на други потребители при вземането на решения за покупки. Анализите могат да бъдат свързани с отделни продукти, магазини и персонал, с цел да се идентифицират постигнати изключителни резултати или области за подобрене.

- *Подобряване удовлетвореността на клиентите*

Използването на технологии като виртуална реалност, разпознаване на образи, изкуствен интелект би подобрило значително избора на продукти и удовлетвореността на клиентите от покупката. Примери в тази област са виртуално пробване на продукти (дрехи, обувки). Lacoste предлага на клиентите си да пробват виртуално обувки, като повече от 30 000 потребители са използвали приложението. Sephora създава приложение за виртуално гримиране, което използва камерата на телефона. През 2003г. шведската фирма IKEA въвежда разпознаването на образи и добавената реалност в каталожните си приложения, с цел потребителите да имат представа как биха изглеждали продуктите в техните домове. Използването на виртуална реалност довежда до поинформиран избор на продукти, по-голяма удовлетвореност на клиентите и по-малко върнати продукти.

- **Откриване на измами (Fraud Detection)**

Актуален в днешно време е проблемът с измамите. Откриването на измами е дейност, която се реализира в партньорство с финансовите организации. Аналитичните инструменти предоставят възможности за наблюдение на транзакциите с кредитни карти в реално време, а технологиите на машинно обучение позволяват да се разработват поведенчески модели, с които да се предсказват измами. Тези анализи в реално време подпомагат разкриването на измами и предварително блокиране на възможните измамни транзакции, с което се предотвратяват големи финансови загуби както за клиента, така и за компанията.

- **Динамично ценообразуване (Dynamic Pricing)**

Външни фактори като сезон, време на деня, трафик, новини за важни събития влияят на търсенето на продукти. Прилагане в реално време на автоматичен анализ на цените на конкуренти (от техни сайтове) и външните данни, които оказват влияние на търсенето, би позволило на бизнеса да адаптира цените незабавно.

Анализът на големи данни може да се използва за сегментиране на клиентите, изграждането на модели за различни типове клиенти, които са готови да платят по-висока цена при различни обстоятелства.

Оптимизация на пространството в магазини (In-store Optimization)

Геопропространствени данни от магазините могат да се използват за проследяване на движението на клиентите в магазините. Анализът на данните от движението на клиентите води до идентифициране на места в магазините, които често и рядко се посещават. Данните от пространствата в магазина и от движението на клиентите могат да бъдат комбинирани с данни за продажбите (история на пазаруването, онлайн дейност и т.н.) с цел създаване на специални навигационни пътеки и списъци за пазаруване.

Заклучение

Големите данни са критична необходимост, с която бизнесът трябва да се справи в процеса на дигитална трансформация. Тъй като изкуственият интелект, машинното обучение и интернет на нещата продължават да бъдат фактори в дигиталната трансформация на бизнеса, компаниите все повече зависят от възможностите за автоматизиран анализ на големи данни от мощни бизнес интелигентни системи. Това неизбежно води до бързо усъвършенстване на технологиите за обработката и анализ на големи данни, за да отговорят на спецификата на данните и изискванията на бизнеса.

Поради своите предимства – ниски разходи и иновативен подход при разработването им, технологиите с отворен код се развиват бързо и постоянно нараства тяхното приложение при реализацията на проекти за използване на големи данни в бизнес дейностите на компаниите. В статията са представени популярни, често използвани технологии с отворен код за обработка на големи данни, които се използват и в решения за големи данни от доставчици като Cloudera, Hortonworks, MapR и др.

Използвана източници

- 2017 Big Data Analytics Market Study, Wisdom of Crowds® Series, December 20, 2017 Dresner Advisory Services, LLC.
- 2018 Big Data Analytics Market Study, Wisdom of Crowds® Series, 2018 Dresner Advisory Services, LLC.
- Big data 2017 – Market Statistics, Use Cases and Trends, Calsoft
- Big Data and the Creative Destruction of Today`s Business Models (2013), ATKearney.
- Brand, K. (2017), What is „Digital Transformation“? – A short and comprehensible definition, <https://kubraconsult.blog/2017/01/18/what-is-digital-transformation%E2%80%8B-a-short-and-comprehensible-definition/>
- Columbus, L. (2018). Big Data Analytics Adoption Soared in the Enterprise in 2018, Forbes.
- Connolly, S. (2012). 7 Key Drivers for the Big Data Market, Hortonworks, <https://hortonworks.com/blog/7-key-drivers-for-the-big-data-market/>
- Formula for Growth: Innovation, Big Data & Analytics, Grocery Manufacturers Association, <https://www.gmaonline.org/issues-policy/collaborating-with-retailers/big-data-analytics/introduction/>
- Gartner IT Glossary <https://www.gartner.com/it-glossary/digitization>, <https://www.gartner.com/it-glossary/digitalization/>
- Ghosh, P. (2018). Business Intelligence and Analytics Trends 2019, <https://www.dataversity.net/business-intelligence-analytics-trends-2019/>
- Gill, N. (2017). Data Ingestion, Processing and Architecture layers for Big Data and IoT, <https://www.xenonstack.com/blog/big-data-engineering/ingestion-processing-big-data-iot-stream/>
- Hurwitz, J., Nugent, A., Halper, F. and Kaufman ,M. (2013). Big Data for Dummies® Published by John Wiley & Sons, Inc.

MapR Industry Guide to Big Data in Retail, <https://mapr.com/>

Mills, T. (2018). Eight Ways Big Data and AI Are Changing the Business World, Forbes.

Murray, S. (2017). IDC Reveals Worldwide Digital Transformation Predictions, <https://www.idc.com/getdoc.jsp?containerId=prUS43188017>

Oracle Big Data Use Cases, <https://www.oracle.com/big-data/guide/big-data-use-cases.html>

Rozados, I. and Tjahjono, B. (2014). Big Data Analytics in Supply Chain Management: Trends and Related Research, 6th International Conference on Operations and Supply Chain Management, Bali, 2014.

СЪДЪРЖАНИЕ

ИНФОРМАЦИОННИ И КОМУНИКАЦИОННИ технологии

ОСНОВНИ ТЕХНОЛОГИИ И ПРИЛОЖНИ НАСОКИ ЗА ДИГИТАЛНА ТРАНСФОРМАЦИЯ НА БИЗНЕСА В СРЕДА НА ГОЛЕМИ ДАННИ Ас. д-р Станимира Йорданова, проф. д-р Камелия Стефанова	5
---	---

МЕНИДЖМЪНТ практика

ИНФРАСТРУКТУРА НА ГОЛЯМО ПРЕДПРИЯТИЕ В БЪЛГАРИЯ – ОСОБЕНОСТИ И ПРОБЛЕМИ Доц. д-р Петя Емилова, докторант Теодора Спасова	25
--	----

ОЦЕНЯВАНЕ НА ФИРМЕНАТА СТРАТЕГИЯ И СТРАТЕГИЧЕСКАТА РЕЗУЛТАТНОСТ НА ФИРМАТА Д-р Ву Мин Ню, докторант Чиджиоке Нуачуку	47
--	----

ЗАСТРАХОВАНЕ и осигуряване

МОРАЛНИТЕ ВРЕДИ В РАБОТАТА НА ЗАСТРАХОВАТЕЛНОТО ДРУЖЕСТВО Проф. д-р ик. н. Христо Драганов, инж. д-р Георги Драганов	64
--	----

СЧЕТОВОДСТВО и одит

ПРОФЕСИОНАЛНАТА КОМПЕТЕНТНОСТ НА ВЪТРЕШНИТЕ ОДИТОРИ В ПУБЛИЧНИЯ СЕКТОР В БЪЛГАРИЯ – ПРЕДИЗВИКАТЕЛСТВА И ВЪЗМОЖНОСТИ Гл. ас. д-р Дияна Иванова	76
---	----

Редколегия на сп. „Бизнес управление“

Красимир Шишманов – главен редактор, Стопанска академия „Д. А. Ценов“ - Свищов

Никола Янков – зам. главен редактор, Стопанска академия „Д. А. Ценов“ - Свищов

Иван Марчевски, Стопанска академия „Д. А. Ценов“ - Свищов

Ирена Емилова, Стопанска академия „Д. А. Ценов“ - Свищов

Любчо Варамезов, Стопанска академия „Д. А. Ценов“ - Свищов

Румен Ерусалимов, Стопанска академия „Д. А. Ценов“ - Свищов

Силвия Костова, Стопанска академия „Д. А. Ценов“ - Свищов

Международна редколегия на сп. „Бизнес управление“

Александру Неделеа – Университет „Стефан Велики“, Сучава, Румъния

Дмитрий Владимирович Чистов, – ФГОБУ ВПО Финансов университет при правителството на руската федерация, Москва, Русия

Йоана Панагорец – Университет Валахия, Търговище, Румъния

Йото Йотов – Драксел университет, Филадельфия, САЩ

Махмуд Ел Батран – Университет Кайро, Кайро, Египет

Наталья Борисовна Голованова – Московски технологически университет, Москва, Русия

Татяна Викторовна Орехова – Донецки национален университет, Виница, Украйна

Тадиа Джукич —Университет в Ниш, Ниш, Сърбия

Ян Тадеуш Дуда – АГН Университет за наука и технологии, Краков, Полша

Виктор Чужиков – Киевски национален икономически университет "Вадим Гетман", Киев, Украйна

Стилов редактор – Анка Танева

Превод на английски език – ст. преп. Цветана Шенкова, ст. преп.

Даниела Стоилова, ст. преп. Иванка Борисова

Превод на руски език – ст. преп. Ирина Иванова

Технически секретар – ас. Живка Тананеева

Отпечатването на списанието за 2019 г. се осъществява с безвъзмездната финансова помощ на Фонд "Научни изследвания" - НП 07/58 по конкурс "Българска научна периодика - 2018 г."

Дадено за печат на 01.03.2019 г., излязло от печат на 13.03.2019 г., формат 70x100/16, тираж 100

© Стопанска академия „Димитър А. Ценов“ – Свищов,
ул. „Ем. Чакъров“ 2, тел.: +359 631 66298

© Академично издателство „Ценов“, Свищов, ул. „Градево“ 24

ISSN 0861 - 6604

БИЗНЕС управление

БИЗНЕС управление 1/2019



ИЗДАНИЕ НА
СТОПАНСКА АКАДЕМИЯ
„Д. А. ЦЕНОВ“ - СВИЦОВ

1/2019

КЪМ ЧИТАТЕЛИТЕ И АВТОРИТЕ НА СПИСАНИЕ „БИЗНЕС УПРАВЛЕНИЕ“

Списание „БИЗНЕС управление“ публикува изследователски статии, методологически и методически разработки и прегледи, рецензии, опит.

1. Обем:

Статии: минимум - 12 страници; максимум – 20 страници;
Прегледи, рецензии, опит: минимум – 5 страници; максимум -10 страници.

2. Депозирание на материалите:

- на хартиен носител и в електронен вид (по E-mail и/или на CD);

3. Технически характеристики:

- изпълнение Word 2003 (минимум);
- размер на страницата - A4, 29-31 реда и 60-65 знака на ред;
- разстояние между редовете 1,5 lines (At least 22 pt);
- шрифт - Times New Roman 14 pt;
- полета - Top - 2.54 см.; Bottom - 2.54 см; Left - 3.17 см; Right - 3.17 см;
- номерация на страницата - долу вдясно;
- текст под линия - размер 10 pt;
- графики и фигури - Word 2003 или Power Point.

4. Оформление:

- наименование на статията, име на автора, научна степен, научно звание - шрифт Times New Roman, 14 pt, с големи букви Bold - центрирано;
- наименование и адрес на местоработата; телефони за контакти и E-mail;
- резюме на български език в обем до 30 реда; ключови думи - от 3 до 5;
- **JEL** класификация на публикациите с икономически характер (<http://ideas.repec.org/j/index.html>);
- основен текст (изложение);
- таблиците, графиките и фигурите се вграждат софтуерно в текста (да позволяват езикова корекция и превод на английски). Цифрите и текстът вътре в тях се изписват с шрифт Times New Roman 12 pt;
- формулите се създават с Equation Editor;

5. Правила за цитиране под линия:

При цитиране да се спазват изискванията на **APA Style (American Psychological Association)**, поместени тук: <https://www.uni-svishtov.bg/?page=page&id=71>

Всеки автор носи отговорност за отстояваните идеи, съдържанието и техническото оформление на своя текст.

6. Контакти:

Главен редактор: тел.: (+359) 631-66-397
Зам.-главен редактор: тел.: (+359) 631-66-299
Стилов редактор: тел.: (+359) 631-66-335
E-mail: zh.tananeeva@uni-svishtov.bg ; bm@uni-svishtov.bg
Адрес: Стопанска академия „Д. А. Ценов“, ул. „Ем. Чакъров“ №2, Свищов, България