

---

---

# تجزیه و تحلیل کلان داده‌ها :

## فناوری‌های مخرب برای تغییر بازی

---

---

مؤلفین:

دکتر آرویند ساتھی

مترجمین:

دکتر جواد وحیدی – دانشگاه علم و صنعت ایران

دکتر رمضان عباس نژادورزی

مهندس زهرا علیجان نژاد بایی



فن‌آوری نوین

---

---

سرشناسه	: ساتھی، آرویند Sathi, Arvind
عنوان و نام پدیدآور	: تجزیہ و تحلیل کلان داده‌ها: فناوری‌های مخرب برای تغییر بازی / مولفین صحیح: مولف اریده ساحا؛ مترجمین جواد وحیدی، رضا عباس نژادورزی، زهرا علیجان نژادبایی.
مشخصات نشر	: بابل: فناوری نوین، ۱۳۹۹.
مشخصات ظاهری	: ۸۳ ص.
شابک	: 978-622-739317-0
وضعیت فهرست نویسی	: فیبا
یادداشت	: عنوان اصلی: Big data analytics: disruptive technologies for changing the game,c2013.
عنوان دیگر	: فناوری‌های مخرب برای تغییر بازی.
موضوع	: داده‌های کلان
موضوع	: Big data
موضوع	: داده‌کاوی
موضوع	: Data mining
موضوع	: هوش تجاری -- داده‌پردازی
موضوع	: Business intelligence -- Data processing
شناسه افزوده	: وحیدی، جواد، ۱۳۴۸ - مترجم
شناسه افزوده	: Vahidi, Javad
شناسه افزوده	: عباس نژاد ورزی، رمضان، ۱۳۴۸ - مترجم
شناسه افزوده	: علیجان نژاد بایی، زهرا، ۱۳۶۶ - مترجم
رده بندی کنگره	: QA۷۶/۹
رده بندی دیویی	: ۰۰۵/۷۴
شماره کتابشناسی ملی	: ۷۳۶۹۳۱۷
وضعیت رکورد	: فیبا

[www.fanavarienovin.net](http://www.fanavarienovin.net)

تلفن: ۰۱۱-۳۲۲۵۶۶۸۷

بابل، کد پستی ۷۳۴۴۸-۴۷۱۶۷

فن آوری نوین

تجزیه و تحلیل کلان داده‌ها: فناوری‌های مخرب برای تغییر بازی  
ترجمه: جواد وحیدی، رمضان عباس نژادورزی، زهرا علیجان نژادبایی.

نوبت چاپ: چاپ اول

سال چاپ: تابستان ۹۹

شمارگان: ۲۰۰

قیمت: ۳۰۰۰۰ تومان

نام چاپخانه و صحافی:

حروف چینی و تایپ: فناوری نوین

شابک: 978-622-7393-17-0

نشانی ناشر: بابل، چهارراه نواب، کاظم بیگی، جنب مسجد منصور کاظم بیگی، طبقه اول

طراح جلد:

تهران، خ اردیبهشت، نبش وحید نظری، پلاک ۱۴۲ تلفکس: ۶۶۴۰۰۱۴۴-۶۶۴۰۰۲۲۰

## درباره نویسنده

دکتر ارورد ساتی معمار ارتباطات جهانی در سطح گسترده در تیم برنامه ریزی اطلاعات در IBM® است. دکتر ساتی دکتری خود را در مدیریت بازرگانی از دانشگاه کارنگی ملون دریافت کرد و تحت نظر برنده جایزه نوبل دکتر هربرت ای سیمون کار کرده است. دکتر ساتی حرفه‌ای دارای تجربه بیش از ۲۰ سال رهبری در مدیریت معماری و تحویل اطلاعات است. تمرکز اصلی خود را در ایجاد چشم‌انداز و نقشه راه برای پیشرفت تجزیه و تحلیل در رهبری مشتریان IBM از ارتباطات راه دور، رسانه‌ها و سرگرمی و سازمان‌های انرژی و آب و برق در سراسر جهان گذاشته است. او چندین کارگاه آموزشی در زمینه ارزیابی داده‌های بزرگ و توسعه نقشه راه انجام داده است. پیش از پیوستن به IBM، دکتر ساتی پیشگام توسعه راه‌حل‌های مبتنی بر دانش برای CRM در گروه Carnegie بود. در Bearing Point، او رهبری توسعه یکپارچه‌سازی سازمانی، MDM، و سیستم‌های پشتیبانی عملیاتی / سیستم‌های پشتیبانی تجاری (OSS / BSS) برای بازار ارتباطات را بر عهده داشت و همچنین راه‌حل‌های افقی برای ارتباطات، خدمات مالی و خدمات عمومی را توسعه داد. در IBM، دکتر ساتی چندین برنامه مدیریت اطلاعات را در MDM، امنیت داده‌ها، اطلاعات کسب و کار را رهبری کرد، و همچنین زمینه‌های مرتبط و نظارت بر معماری را به حساب‌های استراتژیک آی‌بی‌ام ارائه داد. او همچنین تعدادی از کارگاه‌ها و سخنرانی‌ها را در کنفرانس‌های صنعتی در مورد موضوعات فنی از جمله MDM و معماری داده‌ها ارائه کرده است و دو اختراع ثبت شده در پوشش داده‌ها دارد. اولین کتابش، تجربه و تحلیل تجربه مشتری، توسط MC Press در اکتبر ۲۰۱۱ منتشر شد. دکتر ساتی همچنین در تعدادی از کتاب‌های مدیریت اطلاعات نوشته شده توسط سانیل سوارس، مشارکت کرده بود.

## تشکر و سپاس

اول و مهم تر از همه، دوست دارم کار سخت جامعه اطلاعات در ایجاد یک ماده مرجع جهانی را تأیید کنم. به تعداد زیادی از مواد موجود در اینجا اشاره کرده‌ام، از جمله مدل بلوغ تجاری، چارچوب معماری راه‌حل، و تعدادی از مطالعات موردی. دوست دارم از باب کسلی، واین جنسن، و میک فالود برای درک ایده‌ها و سازمان‌دهی مطالب مرجع قدردانی کنم. دوست دارم از تیم دیویس برای تشویق او و ارائه مثال‌های خدمات مالی قدردانی کنم. جف یوناس برای من الهه بخش بود تا با ایده‌ها و عقاید خود آزمایش کنم و قسمت اعظم ستون فقرات این کتاب را برای من فراهم آورد. ایده‌های تکنیکی با کمک بت برانهیل، پل کریستنسن، الیرابت دیل، تامی هارکن، تامی جانستون، ریک هکلر، ابرهارد هکلر، محمد نومن، پیتر هریسون، داریل بی سی، استیو ریگو، و بری ریسون ایجاد شدند. تیم مرکز راه‌حل‌های جهانی دالاس کریسنین، تام اسلید، متی توماس، و جانکی ورا، نظرات ارزشمندی در مورد ایده‌ها ارائه کردند. محال. شاه. امیلین تی جان، لیوا وتورا، ولفگانگ بوسج، استیو تریگ، دون باهش، و جسیکا وایت در این کتاب مولفه‌های ارزشمند تحلیل ارزش کسب‌وکار را ارائه کرده‌اند. من همچنین دوست دارم از تیم مشاوره صنعت بخش ارتباطات کن کیزلیک، تیرک میچلسن، تاشار مهتا، ریچارد لاناها، ریک ماس، و دیوید باک برای تأمین فرصت‌ها، مشتریان و مشارکت در راه‌حل‌های تجزیه و تحلیل کلان داده تشکر کنم. در مرحله بعدی، من دوست دارم از کار عالی تیم تجزیه و تحلیل کسب‌وکار IBM و تیم مشاوره بهینه‌سازی قدردانی کنم. به‌ویژه از آدام گریستینگ، جوزف برید، آنو جین، براس ویس، آپارنا بتگری و جان هلد که ایده‌های پشت‌صحنه کسب‌وکار و موارد کاربردی را از طریق فعالیت‌های مشاوره‌ای خود ارائه دادند. همچنین می‌خواهم از مارک هولست برای مشارکت و ایده‌های نابش در این راه‌حل‌ها تشکر کنم.

از تیم نرم‌افزاری IBM، گروه مطالعات موردی و همچنین نمونه‌های محصول خوبی که را ارائه دادند. می‌خواهم از راجر ری، دن دبرانر و ویهور کومار از کمک آن‌ها در محصول InfoSphere® Streams؛ آرنو منوروران و پاتریک ویلز برای حمایت از اطلاعات Vivisimo®؛ از اندرو کلیبی برای کمک در موتور تجزیه و تحلیل نتازا؛ شانکار ونکاتارامان، گیریش ونکخالیهه، و کارچیک هری هارن برای ینش بزرگشان؛ کلودیو زانچی برای حفظ حریم خصوصی Optim™؛ و مایک زوکر برای SPSS® تشکر کنم. همان‌طور که فرصت‌های تجارت داده‌های بزرگ را مطالعه کردم به‌طور بسیار نزدیک با مشاغل کار کردم. این شامل آنتونی بهان، اش کاناگات، آدری لیرد، باب وایس، کریستین تویفورد، کارمن آلن، دیو دونمایر، داگ هومفریز، دونان گابور، گوتام شاه، گریشی ورما، هارپیندین سینگ مدان، هارس باطنگار، جی پراتوری، جسیکا شاه، جیم هیکس، جاشوا قرآن، جودیت لیست، کیدریک براون، کن باب، لیندزی پردون، مهش دالوی، مورین لایت، نیل ایسفورد، نوربرت هرمان، الیور بچ، پری مک دونالد، فیلیپ اسمولین، پیویش سرال، راوی کوتاری، رندی جورج، راکل کتیگباک، ریچا پاندی، راب اسمیت، رابرت سگات، سام قیین سانکار ویرجارواران، سارا فیلیپوت، استیو کوهن، استیو تیتسل، سومیت چوودوری، سومیت سینگ، ترزا یعقوب،

امدوی ردی، واسکو کوئیرس، ویکاس پاتوری، ون مک کانل و یول آرتی. من از بحث و تفسیر روشنی در درک فرصت‌های تجاری و همچنین شیوه‌های موجود در حال حاضر بزرگ سپاس گزارم.

من از شریل دخترتهتی برای بررسی کتاب و سانیل سوارس برای الهام بخشی به من برای نوشتن کتاب تشکر می‌کنم. گاراو دشیند مقدار قابل توجهی از کار پشت‌صحنه را به من کمک کرد تا کتاب را سازمان‌دهی و تأمین کنم. این نیز الهام گاراو بود برای معرفی نوار کاریکاتور، که در نهایت با همکاری بین ما دو نفر بود. سوزان ویسر کمک قابل توجهی به سازمان‌دهی فرآیند انتشار ارائه کرد. کیتی تاپتون، نگارش ارزشمند و راهنمایی سرمقاله را ارائه داد.

آخرین، اما نه کم‌ترین، از همسر من، دخترم کینجی، پسرانم کوین و کانال برای الهام بخشی، پشتیبانی و کمک برای سرمقاله تشکر می‌کنم.

## پیش گفتار

### توسط باب کسلی

ما شاهد علاقه بی سابقه‌ای به تجزیه و تحلیل داده‌های بزرگ در سراسر جهان هستیم. اعضای برتر اعلام کرده‌اند که خودشان سازمان‌های "Analytics Driven" را هدایت می‌کنند. رهبران کسب و کار و فناوری اطلاعات استفاده از تجزیه و تحلیل داده‌های بزرگ برای به دست آوردن پیشرفت‌های قابل توجه در مدل‌های کسب و کار خود، مشارکت و فرآیندهای کسب و کار آغاز کردند. درحالی‌که تقریباً همه از کلان داده‌ها در سطح ابزار یا محصول صحبت می‌کنند، درحالی‌که تقریباً همه افراد در مورد کلان داده‌ها در سطح ابزار یا محصول صحبت می‌کنند، سازمان‌های موفق در مورد، موارد و تکنیک‌های استفاده از کلان داده‌ها متمرکز می‌شوند که بزرگ‌ترین ارزش کسب و کار را در اختیار دارند. آن‌ها بر روی "کسب و کار" تجزیه و تحلیل داده‌های بزرگ تمرکز می‌کنند. آرویند دیدگاه مشابهی را درباره تجزیه و تحلیل کلان داده‌ها مورداستفاده قرار داده است: فناوری‌های متضاد برای تغییر بازی.

در طی سه سال گذشته، تیم برنامه اطلاعات ما به‌طور گسترده‌ای به سازمان‌ها کمک کرده است تا راهکارهای تحلیل کلان داده‌ها خود را شکل دهند. شروع کسب و کار برای موفقیت هر سازمانی ضروری است. خوشحالم که یک کتاب را که با کسب و کار به‌عنوان تمرکز اصلی و بررسی بهترین شیوه‌های فروش، بازاریابی، خدمات مشتری و مدیریت ریسک، قبل از پیوند دادن آن‌ها به راه‌حل‌ها و معماری‌هایی که آن ممکن می‌سازد، شروع می‌کند آشنا شدم. ما امیدواریم که از این کتاب در مورد پیشرفت بهترین شیوه‌ها و تأثیر آن‌ها در چشم‌انداز رقابتی لذت ببرید. ممکن است گفتگو درست بین رهبران IT و کسب و کار شما را تسهیل کند.

باب Keseley

معاون رئیس جمهور، برنامه اطلاع‌رسانی WW

گروه IBM نرم‌افزار

## پیش‌گفتار

### توسط جف جوناس

این کتاب شامل تعدادی از موارد استفاده کلان داده‌ها، ملاحظات معماری و افزایش فضاها برای بازاریابی در حال ظهور (اجتماعی، جغرافیایی و غیره) است و برخی از مسائل پیچیده اطراف حریم خصوصی داده‌ها را پوشش می‌دهد. فضای مشاهده موجود در سازمان (اطلاعاتی که آن‌ها می‌توانند خود را درون قانون و سیاست قرار دهند) سریع‌تر از توانایی درک آن‌ها رو به رشد است. همان‌طور که سازمان‌ها برای ادامه فعالیت تلاش می‌کنند، آن‌ها مجبورند تجدیدنظر کنند که چه نوع زیرساختی برای مهار داده‌های بزرگ مورد نیاز است.

سازمان‌ها باید در حال پیشروی باشند تا بتوانند تراکنش‌هایی را که اکنون اتفاق می‌افتد حس کنند و پاسخ دهند و باید بتوانند به شدت بر آنچه که مشاهده شده، منعکس شود--- این انعکاس عمیق فعالیت ضروری برای کشف سیگنال ضعیف مرتبط و الگوهای نوظهور است. به دنبال آزمایش‌های اخیر که شامل چگونگی قرارگیری پازل‌های انسانی در کنار هم است، من شاهد بحرانی بودن اتصال شدید از بازتاب عمیق دقیقاً پشت معنای واقعی در زمان و تجزیه و تحلیل پاسخ بودم. در حقیقت، چون حلقه بازخورد سریع‌تر و سریع‌تر می‌شود، این کشف به‌طور چشمگیری افزایش می‌یابد.

سازمان‌هایی که چگونگی ایجاد حس یادگیری سریع برای انجام کاری درباره آن را می‌سنجند، این در حالی اتفاق می‌افتد که بسیار رقابتی‌تر خواهد بود.

جف جوناس

کارمند IBM و دانشمند ارشد

عضو تجزیه و تحلیل IBM

# فهرست مطالب

9	فصل اول: مقدمه.....
10	1-1. حجم .....
11	2-1. تنوع .....
11	3-1. سرعت .....
12	4-1. صحت .....
15	فصل دوم: محرک‌ها برای کلان داده‌ها؟.....
15	1-2. مصرف‌کنندگان پیشرفته .....
19	1-1-2. اتوماسیون .....
20	2-1-2. کسب درآمد .....
23	فصل سوم: برنامه‌های تجزیه و تحلیل کلان داده‌ها .....
23	1-3. مرکز فرماندهی رسانه‌های اجتماعی .....
24	2-3. مرکز دانش محصول .....
27	3-3. مطالعات زیرساختی و عملیاتی .....
28	4-3. انتخاب محصول، طراحی و مهندسی .....
30	5-3. خدمات مبتنی بر مکان .....
32	6-3. قطعه‌سازی میکرو و بهترین آینده .....
36	7-3. تبلیغات آنلاین .....
38	8-3. مدیریت بهبود ریسک .....
40	فصل چهارم: مؤلفه‌های معماری .....
40	1-4. بسترهای نرم‌افزاری پردازش موازی (MPP) .....
43	2-4. تجزیه و تحلیل داده‌ها و گزارش دهی بدون ساختار .....
48	3-4. داده‌های بزرگ و نمای واحد از مشتری / محصول .....
52	4-4. محافظت از حریم خصوصی داده‌ها .....
56	5-4. تحلیلی سازگار و موتورهای تصمیم‌گیری در بی‌درنگ .....
59	فصل پنجم: سکوی تجزیه و تحلیل پیشرفته.....
61	1-5. معماری زمان واقعی برای گفتگوها .....
64	2-5. ارکستراسیون و سنتر با استفاده از موتورهای تجزیه و تحلیل .....
67	3-5. کشف با استفاده از داده‌ها باقی‌مانده .....
69	4-5. استراتژی‌های ادغام .....
71	فصل ششم: اجرای تجزیه و تحلیل.....
71	1-6. تحول، تکاملی یا ترکیبی .....
73	2-6. مدیریت کلان داده‌ها .....
76	3-6. نقاط عطف و سطح رشد .....
81	فصل هفتم: افکار بسته شده.....



تجزیه و تحلیل کلان داده‌ها یک موضوع محبوب است. درحالی‌که همه داستان‌های ارزشیابی جدید دره سلیکون و کمبودهای مهم دانشمندان داده را شنیده‌اند، تعداد قابل توجهی نگرانی وجود دارد، آیا این سرمایه‌گذاری فعلی من در هوش تجاری را از بین خواهد برد یا جایگزین سازمان من خواهد شد؟ چگونه می‌توانم انبار داده و هوش تجاری خود را با داده‌های بزرگ ادغام کنم؟ چگونه می‌توانم شروع کنم، تا بتوانم نتیجه‌ای کسب کنم؟ مهارت‌های موردنیاز چیست؟ چه اتفاقی می‌افتد برای مدیریت داده؟ چگونه می‌توانیم با حریم شخصی داده‌ها مقابله کنیم؟ طی ۹ تا ۱۲ ماه گذشته، من کارگاه‌های زیادی را با پزشکان در این زمینه انجام داده‌ام. من همیشه مجذوب دو منظره‌ای هستم که اغلب در یک اتاق درگیر می‌شوند - کاوشگران چشم‌روشنی که برای به اشتراک گذاشتن داده‌های خود آماده هستند و نگرانی‌هایی که می‌توانند راه‌های ایجاد این مشکل را به وجود آورند. یک تقسیم مشابه در میان مصرف‌کنندگان وجود دارد. همان‌طور که در هر زمینه‌ی جدید اجرای کلان داده نیاز به یک تعادل ظریف بین دو دیدگاه و یک معماری قوی دارد که می‌تواند نگرانی‌های مختلف را تطبیق دهد. برخلاف بسیاری از وبلاگ‌ها و کتاب‌های بزرگ تجزیه و تحلیل کلان داده‌ها که اصول و پایه‌های فناوری را پوشش می‌دهند، این کتاب از دیدگاه یک پزشک برخوردار است. این موارد استفاده برای تجزیه و تحلیل کلان داده‌ها، اجزای مهندسی آن و چگونگی ادغام کلان داده‌ها با فرآیندها و سیستم‌های تجاری را مشخص می‌کند. در انجام این کار، سرمایه‌گذاری‌های کلان در انبار داده‌ها و هوش تجاری رجوع می‌کند و هر دو روش تکاملی و انقلابی و همچنین ترکیبی از حرکت به جلو دنیای جدید کلان داده‌ها را نشان می‌دهد. این موضوع به موضوعاتی جدی درباره‌ی حفظ حریم شخصی داده‌ها و حاکمیت شرکت‌ها پرداخته است و چگونه باید در اجرای برنامه‌های کلان داده‌ها برای محافظت از داده‌های مان، حریم خصوصی مشتریان و محصولات مان مراقبت کرد. بنابراین کلان داده چیست؟ دو منبع متداولی از اطلاعات گروه در زیر پرچم کلان داده‌ها وجود دارد. اولاً یک مقدار مناسبی از اطلاعات درون شرکت داریم که به لطف اتوماسیون و دسترسی به‌طور فزاینده‌ای به اشتراک گذاشته شده است. این شامل ایمیل‌ها، لگاریتم‌های مربوط به سیستم عامل، وبلاگ‌ها، اسناد تصویرسازی پی‌دی‌اف، رویدادهای فرایند کسب و کار و هرگونه اطلاعات ساختاری، غیر ساختاری یا نیمه ساختاری موجود در داخل سازمان می‌شود. دوماً اطلاعات بیش‌تری در خارج از

سازمان می‌بینیم. بعضی از آن‌ها در دسترس عموم آزاد هستند، برخی بر اساس اشتراکی پرداخت می‌شوند و بقیه به‌طور انتخابی برای شرکای تجاری یا مشتریان خاص انتخاب می‌شوند. این شامل اطلاعات موجود در سایت‌های رسانه‌های اجتماعی، ادبیات آزادانه توسط رقبا، سلسله‌مراتب سازمان‌های مشتریان، شرکت‌ها، نکات مفید در دسترس شخص ثالث و شکایات مشتریان در سایت‌های نظارتی منتشرشده است. بسیاری از سازمان‌ها در تلاش هستند تا مشتریان را برای ایجاد اطلاعات جدید تشویق کنند. به‌عنوان مثال، foursquare ([www.foursquare.com](http://www.foursquare.com)) به مجموعه‌ای از کسب‌وکارهای تبلیغاتی از طریق foursquare تشویق می‌کند که بازدیدکنندگان سند بگیرند. این برای من یک فرصت برای هر بازدید فراهم می‌کند و اگر بیش‌تر بازدیدکننده‌ای از یک مکان کسب‌وکار خاص هستم به‌عنوان شهردار پاداش می‌دهد. به‌عنوان مثال، وقتی که من به مکان موردعلاقه‌ام در نزدیکی سوشی توکیو جوزه می‌روم اجازه می‌دهم foursquare در مورد ملاقاتم بدانند و امتیازات جایزه را جمع‌آوری کند. احتمالاً foursquare، توکیو جوزه تمام رستوران‌های رقابتی سوشی می‌توانند از این اطلاعات برای جذب توجه من در فرصت‌های بعد از غذا استفاده کنند. Sunil soares پنج نوع داده‌ی بزرگ را شناسایی کرده است: وب و رسانه‌های اجتماعی، ماشین به ماشین (M2M)، معامله‌ی کلان داده‌ها، بیومتریک و تولیدی انسان.

بعضی از نمونه‌های کلان داده‌ها که در این کتاب استفاده خواهیم کرد:

- ❖ متن رسانه‌های اجتماعی
- ❖ مکان‌های تلفن همراه
- ❖ اطلاعات کلیک کانال از جعبه‌ی تنظیمات
- ❖ مرور و جست‌وجو در وب
- ❖ راهنماهای محصول
- ❖ رویدادهای شبکه‌ی ارتباطات
- ❖ سوابق جزئیات تماس (CDRs)
- ❖ شناسایی فرکانس رادیویی (RFID)
- ❖ نقشه‌ها
- ❖ الگوهای ترافیکی
- ❖ اطلاعات آب‌وهوا
- ❖ لگاریتم سیستم اصلی

## مقدمه ۱۱

چرا کلان داده با هر داده‌ی دیگری که در گذشته به آن پرداخته‌ایم متفاوت است؟ چهار V وجود دارد که این داده‌ها را مشخص می‌کند: حجم، سرعت، تنوع و صحت. بعضی از تحلیلگران Vهای دیگری را به این لیست اضافه کرده‌اند، اما برای اهداف این کتاب، من در مورد چهار V در اینجا توضیح خواهم داد.

### ۱-۱. حجم

اکثر سازمان‌ها با افزایش اندازه‌ی داده‌های پایگاه خود در حال مبارزه با سونامی کلان داده‌هایی هستند که به انبارهای داده ضربه می‌زدند. طبق مجله‌ی فورچون، ما پنج اگزابایت داده‌ی دیجیتال را در زمان ضبط تا سال ۲۰۰۳ ایجاد کردیم. در سال ۲۰۱۱ همان مقدار داده‌ها در دو روز ایجاد شد. انتظار می‌رود تا سال ۲۰۱۳ این دوره‌ی زمانی تا ۱۰ دقیقه کاهش یابد. دهه‌ی پیش سازمان‌ها معمولاً ذخیره‌سازی داده‌ها را برای زیرساخت‌های تحلیلی در ترابایت شمارش کردند. آن‌ها اکنون به برنامه‌هایی نیاز دارند که در حجم پتابایت ذخیره کنند. این داده‌ها زیرساخت تجزیه و تحلیل در تعدادی از صنایع را تضعیف می‌کند. برای یک ارائه‌دهنده‌ی خدمات ارتباطی (CSP) با ۱۰۰ میلیون مشتری می‌تواند اطلاعات روزانه‌ی مربوط به مکان حدود ۵۰ ترابایت را داشته باشد که در صورت نگه‌داری برای ۱۰۰ روز حدود ۵ پتابایت را اشغال می‌کنند. در بحث‌هایی با یک شرکت کابل متوجه شدم که اکثر داده‌های شبکه‌ی خود را در پایان از بین می‌برند، از آنجاکه آن‌ها ظرفیت ذخیره‌ی آن را ندارند اگرچه تنظیم‌کننده‌ها اکثر CPUها و اپراتورهای کابلی را برای ذخیره اطلاعات مربوط به جزئیات تماس و اطلاعات مربوط به استفاده از آن دارند. برای ۱۰۰ میلیون CDR مشترک، CPSها به راحتی می‌توانند بیش از ۵ میلیارد پرونده در روز داشته باشند. از سال ۲۰۱۰ AT و T در پایگاه داده‌ی خود دارای ۱۹۳ تریلیون CDR بوده است.

### ۲-۱. سرعت

دو جنبه برای سرعت وجود دارد. یکی از آن نشان‌دهنده‌ی توان پردازش اطلاعات است و دیگری نشان‌دهنده‌ی تأخیر است. اجازه دهید با توانایی شروع کنیم که نشان‌دهنده‌ی داده‌های در حال حرکت در لوله‌هاست. میزان داده‌های جهانی تلفن همراه با رشد ۷۸ درصدی همراه است و انتظار می‌رود که در سال ۲۰۱۶ به ۱۰.۸ اگزابایت در هر ماه برسد. زیرا مصرف‌کنندگان تصاویر و فیلم‌های بیش‌تری را به اشتراک می‌گذارند. برای تجزیه و تحلیل این داده‌ها، زیرساخت تجزیه و تحلیل شرکت به دنبال لوله‌های بزرگ‌تر و پردازش گسترده موازی است. تأخیر یکی دیگر از راه‌های اندازه‌گیری سرعت است. تجزیه و تحلیل به‌عنوان یک محیط ذخیره و گزارش که در آن گزارشات به‌طور معمول حاوی اطلاعات از دیروز بود به‌طور عمده به‌عنوان D<sub>1</sub> نشان داده شده است. اکنون، تجزیه و تحلیل به‌طور فزاینده‌ای در فرآیندهای تجاری با استفاده از داده‌های در حال حرکت با کاهش زمان

تأخیر درج می‌شود. به‌عنوان مثال، (www.turn.com) Turn در حال انجام تجزیه و تحلیل خود در ۱۰ میلی ثانیه برای قرار دادن تبلیغات در سیستم عامل‌های تبلیغات آنلاین است.

### ۳-۱. نوع

در دهه ۱۹۹۰، با پیشرفت سریع فناوری انبار داده، تلاش اولیه ایجاد مدل‌های متا بود تا تمام داده‌ها را در یک قالب استاندارد نشان دهند. داده‌ها از منابع مختلف جمع‌آوری شده و با استفاده از ETL تبدیل شده‌اند (استخراج، تبدیل، بار) و یا ELT (داده‌ها را استخراج کرده و آن را در انبار بارگذاری می‌کنید، و سپس آن را درون انبار تبدیل می‌کنید). فرض اساسی محتوای محدود و محتوای ساختاری بود. کلان داده به‌طور چشم‌گیری افق‌های ما را افزایش داده است که با ادغام داده‌های جدید و تکنولوژی‌های تجزیه و تحلیل انجام شده است. تعدادی از راه‌حل‌های تجزیه و تحلیل مرکز تماس به دنبال تجزیه و تحلیل مکالمات مرکز تماس و ارتباط آن‌ها با ایمیل‌ها، مشکلات بلیط‌ها و وبلاگ‌های اجتماعی هستند. علاوه بر داده‌های ساخت یافته، داده‌های منبع شامل متن، صدا و ویدیوی بدون ساختار است. تعدادی از برنامه‌های جمع‌آوری داده‌ها از ایمیل، اسناد و یا وبلاگ‌هاست. به‌عنوان مثال ضریب محاسبات تجزیه و تحلیل سفارش را برای سفارشات آنلاین فراهم می‌کند (برای اطلاعات بیشتر به [www.slice.com](http://www.slice.com) مراجعه کنید). ایمیل‌های تجزیه و تحلیل شده، و اطلاعات داده‌های خام را جستجو می‌کنند.

انواع سازمان‌ها، بلیط‌های هواپیمایی، خرید آنلاین کتاب‌فروشی‌ها، رسیدهای دریافت موسیقی، بلیط پارکینگ شهر و یا هر چیزی که می‌توانید خریداری کنید و هزینه‌ی آن را در ایمیل خود مشاهده کنید.

چگونه می‌توان این اطلاعات را به یک کاتالوگ محصول تبدیل و خرید را تجزیه و تحلیل کرد؟

یکی دیگر از تکنولوژی‌های فعال، پلت‌فرم IBM InfoSphere Streams است که با منابع مختلفی برای تجزیه و تحلیل و تصمیم‌گیری در زمان بلادرنگ، از جمله در ابزار پزشکی برای تجزیه و تحلیل نوزادان، لرزه‌نگاری داده‌ها، CDR، رویدادهای شبکه، برچسب‌های RFID، الگوهای ترافیک، داده‌های آب‌وهوا، سیستم عامل اصلی و صدا در بسیاری از زبان‌ها و ویدیوها است.

### ۴-۱. صحت

برخلاف داده‌های داخلی که با دقت اداره می‌شوند، اکثر کلان داده‌ها منابع خارج از کنترل ما هستند. بنابراین، از صحت مشکلات و دقت قابل توجهی رنج می‌برند. صحت نشان‌دهنده‌ی اعتبار منبع داده و همچنین مناسب بودن داده‌ها برای هدف مخاطبان است. اجازه دهید با اعتبار منبع شروع کنیم. اگر یک سازمان برای جمع‌آوری اطلاعات مربوط به محصول از طرف اشخاص ثالث و ارائه آن به کارمندان مرکز تماسش از مشتریان نمایشگاه پشتیبانی کند،

## مقدمه ۱۳

داده‌ها باید برای دقت و اعتبار منبع نمایش داده شود. در غیر این صورت، مراکز تماس می‌توانند پیشنهادات رقابتی ارائه دهند که ممکن است پیشنهادات را کنار بگذارند و فرصت‌های درآمد را کاهش دهند.

تعداد زیادی از پاسخ‌های رسانه‌های اجتماعی به کمپین‌ها می‌تواند از تعداد کمی از کارکنان ناراضی گذشته یا در رقابت افراد مشغول به کار برای ارسال نظرات منفی باشد. به عنوان مثال، فرض می‌کنیم که like در یک محصول نشان‌دهنده‌ی مشتریان راضی است اگر like توسط یک شخص ثالث قرار داده شود چه؟ همچنین باید در مورد مناسب بودن مخاطب فکر کنیم و این که چگونه می‌توان حقیقت را با یک مخاطب خاص به اشتراک گذاشت. اطمینان از داده‌های ایجادشده در یک سازمان حداقل می‌تواند رضایت‌بخش باشد. باین حال، برخی از داده‌های داخلی امکان دارد برای ارتباط گسترده‌تر در دسترس نباشند. به عنوان مثال، اگر خدمات مشتری در مورد نواقص محصول همان‌طور که در نقاط تماس مشتری مشاهده می‌شود، مهندسی شده است، این داده‌ها باید به صورت انتخابی و بر اساس نیاز به اشتراک گذاشته شوند. سایر داده‌ها فقط ممکن است با مشتریانی که قراردادهای معتبر یا پیش‌نیازهای دیگری دارند، به اشتراک گذاشته شوند.

طی یک سال گذشته، از تیم دستور کار اطلاعات خواسته‌شده است که تعدادی از کارگاه‌های تجزیه و تحلیل کلان داده‌ها را برگزار کنند.

سه سؤال رایج عبارت‌اند از:

۱. کلان داده‌ها چه هستند؟ دیگران با آن چه کاری انجام می‌دهند؟
۲. چگونه یک برنامه‌ی استراتژیک برای تجزیه و تحلیل کلان داده‌ها در پاسخ به درخواست مدیریت ایجاد کنیم؟

۳. چگونه کلان داده‌ها تجزیه و تحلیل سازمان و معماری ما را تغییر می‌دهد؟

اکثر موارد موجود در این کتاب پاسخ به این سؤالات است. این کتاب سه دیدگاه را در تجزیه و تحلیل کلان داده‌ها فراهم می‌کند.

اول، چرا تجزیه و تحلیل کلان داده بسیار مهم است و با چه چیزی می‌توانیم انجام دهیم؟ این کتاب بزرگ‌ترین روندهای رشد اطلاعات کلان را نشان می‌دهد که در موارد معمولی از تجزیه و تحلیل کلان داده استفاده شده است، جایی که سازمان‌های برجسته در حال حاضر از مزایای عمده‌ی آن دیدن می‌کنند.

دوم، این کتاب مؤلفه‌های اصلی تجزیه و تحلیل کلان داده‌ها را نشان می‌دهد و یک معماری یکپارچه AAP را معرفی می‌کند که تجزیه و تحلیل کلان داده‌ها را با بقیه‌ی زیرساخت‌های تحلیلی ترکیب می‌کند و با فرایندهای

تجاری ادغام می‌شود. این نشان می‌دهد که چگونه این اجزا در AAP باهم کار می‌کنند تا یک موتور یکپارچه را ارائه دهند که می‌تواند کلان داده‌ها را با انبار داده‌های سنتی و اطلاعات کسب و کار ترکیب کند تا یک راه حل کلی فراهم کند.

سوم، کتاب نگاهی اجمالی به نگرانی‌های اجرا و چگونگی برخورد با آن‌ها را فراهم می‌کند. چگونه می‌توان یک نقشه راه ایجاد کرد و برنامه‌های اصلی آزمایشی را برای جمع‌آوری حرکت و پافشاری برای ایجاد چشم‌انداز تغییر بازی پیاده‌سازی کرد؟ وقتی داده‌های اصلی ممکن است از نظر کیفیت یا محدودیت‌های حریم خصوصی متفاوت باشند، چگونه می‌توانیم حاکمیت را برای این داده‌ها فراهم کنیم؟

فیل بزرگ موجود در اتاق، حریم شخصی داده‌ها است. اعتراف می‌کنم که من در مورد حریم شخصی داده‌ها موضع نگرفته‌ام، و نه پیش‌بینی کرده‌ام که چگونه جهان با آن برخورد خواهد کرد. این یک موضوع در حال تکامل است، با بسیاری از مشکلات، اختلافات جغرافیایی، و عواقب ناشناخته. با این حال، من تعدادی از مناطق مهم را برای بررسی بیشتر، و همچنین تعدادی از اجزای مورد نیاز، صرف نظر از موقعیت گرفته شده، شرح داده‌ام.

من بشدت بر کار شخصی خود برای تصاویر مفاهیم مورد بحث در این کتاب تکیه کرده‌ام. در نتیجه، بیش تر نمونه‌ها به سمت CSPها، تبلیغات و صنایع خرده‌فروشی تقسیم می‌شوند. این به این معنی نیست که این بسته صنایع را هدایت می‌کند یا این که سایر صنایع فرصت‌های کلان داده را ندارند. برعکس، در بسیاری از صنایع نمونه‌های زیادی پیدا کرده‌ایم.

## محرک‌ها برای کلان داده‌ها؟

در حال افزایش سرعت ایجاد کلان داده‌ها هستیم. این فصل نیروهای پشت این سونامی کلان داده را بررسی می‌کند. سه عامل کمک‌کننده وجود دارد: مصرف‌کنندگان، اتوماسیون و کسب درآمد. بیش از هر یک از این عوامل کمک‌کننده، تعامل آن‌هاست که سرعت ایجاد کلان داده را افزایش می‌دهد. با افزایش اتوماسیون، آسان‌تر است که ایجاد کلان داده‌ها و فرصت‌های استفاده از آن را به مصرف‌کنندگان ارائه دهیم و روند کسب درآمد، به‌طور فزاینده‌ای یک بازار مؤثر برای کلان داده‌ها را فراهم می‌آورد.

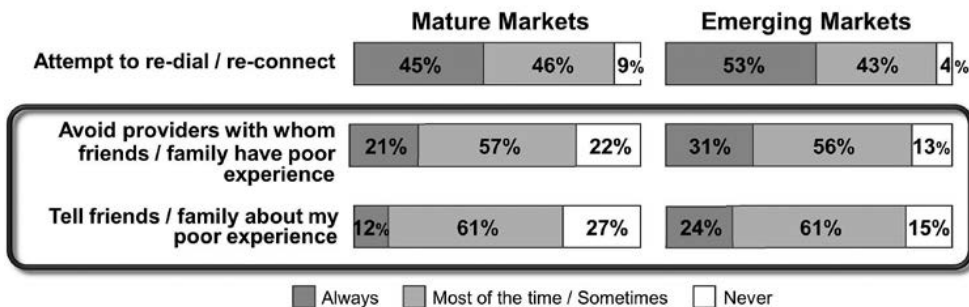
### ۱-۲. مصرف‌کنندگان پیشرفته

افزایش سطح اطلاعات و ابزارهای مرتبط باعث ایجاد یک نسل جدید از مصرف‌کنندگان سطح بالا شده است. این مصرف‌کنندگان تحلیل‌گران بسیار متفاوتی هستند، در استفاده از آمار داناتر، و به‌مراتب بیش‌تر مرتبط هستند، با استفاده از رسانه‌های اجتماعی به‌سرعت به جمع‌آوری و تطبیق نظرات دیگران می‌پردازند. در جهان سرشار از پیام‌های بازاریابی زندگی می‌کنیم. درحالی‌که اکثر بازاریابی‌ها هنوز هم با استفاده از روزنامه، مجله، شبکه تلویزیونی، رادیو و تبلیغات صفحه‌نمایش، حتی در رسانه‌های متعارف، به تدریج پخش تنگاتنگ برجسته‌تر می‌شود. این در نوشته‌های تبلیغاتی محلی در مجلات دیده می‌شود، قرار دادن آگهی‌های تبلیغاتی محدود با استفاده از جعبه‌های تنظیم‌شده، و استفاده از اطلاعات جاده‌ای برای تغییر تبلیغات صفحه‌نمایش. دنیای اینترنت می‌تواند بسیار شخصی شود. موتورهای جستجو، سایت‌های شبکه‌های اجتماعی و صفحات زرد الکترونیکی، تبلیغات ویژه‌ای را برای یک فرد یا یک بخش کوچک ایجاد می‌کنند. از کوکی‌های اینترنتی به‌طور فزاینده‌ای برای ردیابی رفتار کاربر و متناسب‌سازی محتوای مبتنی بر این رفتار استفاده می‌شود. ایمیل‌ها و پیام‌های متنی به‌سرعت منجر به افزایش تعاملات بین فردی شد. ارتباطات نه‌تنها توسط بازاریابان، بلکه همچنین با اشخاص ثالث و دوستان آغاز شد. ارتباطات به تابلوهای اعلانات، گروه‌های گفتگو و رسانه‌های اجتماعی گسترش یافته است به ما اجازه می‌دهد که درباره اهداف، ترس، انتظارات و ناامیدی‌های ما با گروه‌های اجتماعی کوچک و بزرگ گفتگو کنیم. برخلاف ایمیل و متن، مکالمات در وب برای خواندن دیگران، هم‌اکنون و یا زمانی دیگر امکان‌پذیر است.

تاکنون، فقط با اشکال مختلف ارتباطات روبرو شده‌ایم. مجموعه‌های بعدی منابع، اطلاعات را از بیش از یک رسانه ترکیب می‌کنند. به‌عنوان مثال، مکالمات فیس‌بوک شامل تعدادی از رسانه‌ها، از جمله متن، کلیپ‌های صوتی، عکس‌ها و ویدیو است. جهان دوم و واقعیت جایگزین، در حال تبدیل شدن به راه‌های جالبی برای آزمایش ایده‌های محصولات در یک جهان شبیه‌سازی شده است، که با استفاده از آن محصولات، می‌توان آزمایش کرد.

اغلب به کارشناسانی نیاز داریم تا به در مرتب کردن ویژگی‌های محصول و چگونگی ارتباط آن‌ها کمک کنند. امروزه متخصصین زیادی برای کمک به ما در استفاده، تشخیص کیفیت، قیمت‌گذاری و اطلاعات مربوط به ارزش محصولات در دسترس هستند. تعدادی از بازاریابان برنامه‌های مشاور یا سفیر را با استفاده از رسانه‌های اجتماعی تشویق می‌کنند. این مشتریان انتخاب شده پیش‌نمایش محصولات جدید را دریافت می‌کنند و به‌طور فعال در ارزیابی و ترویج محصولات جدید شرکت می‌کنند. در پایان روز، افرادی که می‌شناسیم و به آن‌ها اعتماد داریم، تصمیمات ما را تغییر می‌دهند. این بزرگ‌ترین سهم شبکه‌های اجتماعی است. آن‌ها مصرف‌کنندگان را دور هم جمع کرده‌اند به‌طوری که هم اکنون به اشتراک گذاشتن تجارب مشتری بسیار بیشتر از گذشته شده است.

چگونه یک مصرف‌کننده با کیفیت ضعیف خدمات برخورد می‌کند؟ شکل ۱-۲ رفتارهای معمول را در افراد بالغ نشان می‌دهد و بازارهای در حال ظهور به‌عنوان مطالعات انجام‌شده توسط IBM Global Telecom Consumer Survey با حجم نمونه ۱۰،۱۷۷،۷ انجام شده است. در این نظرسنجی، ۷۸ درصد از مصرف‌کنندگان بررسی شده در بازارهای بالغ گفتند که آن‌ها از ارائه‌دهندگان که دوستان یا خانواده آن‌ها تجربه بد دارند اجتناب می‌کنند. این درصد در بازارهای در حال رشد حتی بالاتر بود (۸۷ درصد). در پاسخ به یک سؤال مرتبط، شرکت‌کنندگان در این نظرسنجی گفتند که آن‌ها با دوستان و خانواده درباره تجربه ضعیف صحبت می‌کنند (۷۳ درصد در بازارهای بالغ و ۸۵ درصد در بازارهای در حال رشد). این اعداد با هم نشان‌دهنده نفوذ قوی شبکه اجتماعی در رفتار خرید است. این درصد بسیار مهمی است و در حال حاضر به‌طور گسترده‌ای توسط سایت‌های و رسانه‌های اجتماعی افزایش می‌یابد (برای مثال، دکمه «Like» در فیس‌بوک قرار داده شده است). همان نظرسنجی همچنین دریافت که سه منبع برای اطلاعات توصیه‌شده اینترنت، توصیه‌های خانواده / دوستان و رسانه‌های اجتماعی ترجیح داده شده است.



منبع: IBM Global Telecom Consumer Survey, Global N = ۱۰۱۷۷; Mature Countries N = ۷۸۷۵

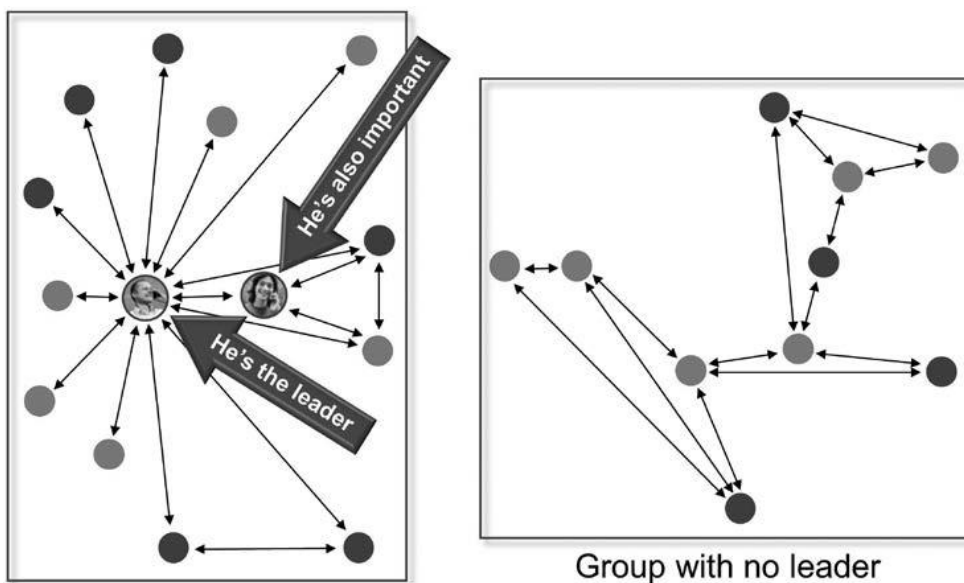
شکل ۱-۲. رفتارهایی در پاسخ به تجربه کیفیت پایین خدمات.



## ۱۷ محرک‌ها برای کلان داده‌ها

در هر گروه، رهبرانی وجود دارند. این‌ها کسانی هستند که تغییرات را از یک نام تجاری به نام دیگری هدایت می‌کنند. رهبران معمولاً مجموعه‌ای طرفدار و پیرو دارند. هنگامی که یک رهبر یک نام تجاری را عوض می‌کند، این احتمال را برای اعضای گروه اجتماعی نیز افزایش می‌دهد. این رهبران چه کسانی هستند؟ آیا می‌توانیم آن‌ها را شناسایی کنیم؟ چگونه می‌توانیم بازاریابی خود را با این رهبران هدایت کنیم؟

در هر ارتباطی، رهبران همیشه مرکزیت دارند (به شکل ۲-۲ نگاه کنید). آن‌ها اغلب به تعداد بیشتری از "پیروان" متصل می‌شوند، بعضی از آن‌ها نیز می‌توانند رهبر باشند. با توجه به شکل، رهبران فلش-های ارتباطی بیشتری دارند که در مقایسه با دیگران هم از آن‌ها شروع می‌شود و هم پایان می‌دهند.



شکل ۲-۲. رهبران در یک شبکه ارتباطی.

چگونه رهبران را شناسایی می‌کنیم؟ تحقیقات IBM یک سری آزمایش با CSP انجام داد. سوابق جزئیات تماس، که اطلاعات مربوط به شخص A را که فرد B را فراخوانی می‌کند، در اختیار شما قرار می‌دهد. با تلفیق اطلاعات تماس و انتزاع شبکه‌های ارتباطی، شبکه‌های ارتباطی را در میان افراد کشف کردیم. همچنین از اطلاعات محرمانه مشتریان استفاده کردیم تا در میان رهبران برای پیروان بعدی همبستگی وجود داشته باشد. در اینجا به برخی از نکات برجسته‌ای در یکی از تجربیاتی که به انجام آن کمک کردم، اشاره می‌کنم:

❖ رهبران در مقایسه با افراد غیر رهبر ۱,۲ برابر بیشتر احتمال دارد که خفه شوند.

- ❖ دو نوع رهبر وجود دارد: توزیع رهبران که از طریق تماس‌های خروجی به گروه خود متصل شده‌اند، و رهبران مرجع که از طریق بخش بیشتری از تماس‌های دریافتی به آن‌ها وصل شده‌اند.
- ❖ هنگامی که یک رهبر پخش‌کننده متفرق شد، چارچوب اضافی ۲۸,۵ برابر بیشتر بود. هنگامی که یک رهبر اقتدار به هم می‌ریزد، احتمال ایجاد چارچوب اضافی ۱۹,۹ برابر بیشتر است.
- ❖ به‌طور معمول، زمان بسیار محدودی بین پیرو یاز رهبران و پیروی از پیروان وجود داشت. گروه‌های اجتماعی را می‌توان از هر نوع ایمیل ارتباطی، متن پیامکی، تماس، دوستی فیس‌بوک و غیره استنباط کرد. دیدن آمارهای قوی مرتبط با تأثیر رهبران بر گروه جالب است. روش‌های زیادی برای استفاده از شبکه‌های اجتماعی برای تأثیرگذاری در خرید و استفاده مجدد وجود دارد:

❖ **مطالعه تجربه مصرف‌کنندگان:** مقدار مناسبی از این داده‌ها بدون ساختار است. با تجزیه و تحلیل احساساتی متن، شدت، خوانندگان، وبلاگ‌های مرتبط، ارجاع‌ها و سایر اطلاعات، می‌توانیم داده‌ها را با توجه توانایی‌های مثبت و منفی و تأثیر آن‌ها بر پایه مشتری سازمان‌دهی کنیم.

❖ **سازمان‌دهی تجربه مشتری:** می‌توانیم بررسی‌ها را به یک خریدار آینده‌نگر ارائه دهیم تا بتوانیم چگونگی ارزیابی محصول را بررسی کنیم.

❖ **تأثیر شبکه‌های اجتماعی:** می‌توانیم اساس بازاریابی، تغییرات محصول، دستورالعمل‌های شرکت و حمایت‌های مشهور در شبکه‌های اجتماعی را ارائه دهیم، به‌طوری که رسانه‌های اجتماعی می‌توانند به افزایش و تأثیر سرریز کمک کنند.

❖ **بازخورد برای محصولات، عملیات، و یا بازاریابی:** با استفاده از اطلاعات تولیدشده توسط رسانه‌های اجتماعی، می‌توانیم به سرعت تغییرات در ترکیب محصول و بازاریابی برای بهبود به مشتریان را ارائه دهیم.

جامعه همیشه در روند ارزیابی نقش مهمی ایفا کرده است. باین حال، اینترنت و شبکه‌های اجتماعی به‌طور اساسی دسترسی ما به اطلاعات را تغییر داده‌اند. من ممکن است "Like" محصول را در فیس‌بوک انتخاب کنم، که شبکه هم‌اکنون دسترسی سریع به این عمل دارد. اگر یک رستوران ارزش پول خود را در نظر بگیرد، Yelp می‌تواند کمک کند که این واقعیت در سراسر جهان پخش شود. اگر از سرویس‌های Csp تلفن همراه جدید راضی نباشیم می‌توانیم شکایت آن را به همه بگوییم.

## ۲-۲. اتوماسیون

پاسخ‌های صوتی تعاملی (IVR)، کیوسک‌ها، دستگاه‌های تلفن همراه، ایمیل، چت، وب‌سایت‌های شرکت‌های بزرگ، برنامه‌های شخص ثالث و شبکه‌های اجتماعی مقدار مناسبی از اطلاعات و وقایع را در مورد مشتریان ایجاد کرده‌اند. علاوه بر این، تعاملات مشتری از طریق رسانه‌های سنتی مانند مراکز تماس می‌تواند اکنون تحلیل و سازمان‌دهی شود. بزرگ‌ترین توانایی ما برای تغییر تجربه مشتری، استفاده از سیاست‌ها، روش‌ها و شخصی‌سازی نرم‌افزار است، ایجاد خدمات، خود به‌طور فزاینده‌ای مشتری خوب را افزایش می‌دهد. فروش و بازاریابی در ده سال گذشته از ابزار اتوماسیون مبتنی بر اینترنت، بزرگ‌ترین و بیشترین ترقی را در ابزار دقیق داشته است.

مرور، خرید، سفارش و خدمات مشتری در وب نه‌تنها کنترل بسیار زیادی را برای کاربران فراهم کرده است، بلکه یک سیل عظیمی از اطلاعات را به سازمان بازاریابی محصول و فروش در درک رفتار خریدار ایجاد کرده است. هر دنباله‌ای از کلیک‌های وب را می‌توان جمع‌آوری و رضایت مشتری، اختلاف‌نظر و یا نقض مستقیم را تجزیه و تحلیل کرد. اطلاعات بیشتر نیز در مورد توالی تصمیم‌گیری به دست می‌آید. خدمات خود را از طریق وسیله‌های مختلف از جمله IVRs، کیوسک، دستگاه‌های دستی، و بسیاری دیگر می‌توان ارائه کرد هر یک از این ابزارهای ارتباطی الکترونیکی مانند یک استخر عظیم از مطالعات زمان و حرکت عمل می‌کنند. در مورد چگونگی تعداد مراحل مشتریان اطلاعات داریم، چند محصول مقایسه شده است، و تمرکز روی چه ویژگی‌هایی بوده است، مانند قیمت‌ها، ویژگی‌ها، مقایسه برندها، توصیه‌ها، نقص‌ها و غیره. تولیدکنندگان مقادیر عظیمی از داده‌های سرویس‌های خود و حسگرهای الکترونیکی متصل به محصولات را به دست آورده‌اند. اگر برای تماشای تلویزیون از یک جعبه تنظیم دوطرفه استفاده کنم، تأمین‌کننده، دسترسی فوری به چگونگی گشتنم در کانال را دارد. آیا با شروع تبلیغات، کانال را تغییر دادم؟ وقتی صدای زنگ شروع به پخش شد، مقدار صدا را افزایش یا کاهش دادم؟ اگر از اینترنت برای خرید یک محصول استفاده می‌کنم، جریان کلیکم قابل تجزیه و تحلیل است و برای مطالعه رفتار خرید استفاده می‌شود. چند تا محصول را نگاه کردم؟ آیا وقتی محصولی را دیدم، توضیح محصول یا قیمت آن را مشاهده کردم؟ این مجموعه غنی از داده‌ها ما را قادر می‌سازد تا تجربه مشتری را در جزئیات دقیق‌تر تحلیل کنیم.

منابع اطلاعاتی از جمله تعاملات خودآموزی چیست؟

❖ **محصول:** هرچه محصولات به‌صورت فزاینده الکترونیکی می‌شوند، داده‌های ارزشمند زیادی را در رابطه با استفاده و کیفیت آن محصول ارائه می‌دهند. در بسیاری از موارد، تأمین‌کنندگان همچنین می‌توانند اطلاعاتی را در مورد زمینه‌ای که در آن از یک محصول استفاده شده است، جمع‌آوری کنند.