فصلنامه تحقیقات جدید در علوم انسانی
Human Sciences Research Journal
دوره جدید، شماره ۲۶، بهار ۱۳۹۹، صص ۱۳۰-۱۰۹ **New Period, No 26, 2020, P 109-130**
شماره شاپا (۲۴۷۶-۷۰۱۸) **ISSN (2476-7018)**

بررسی کلان داده‌ها در شبکه‌های اجتماعی

الهام حیدری
دانشجوی کارشناسی ارشد علم اطلاعات و دانش‌شناسی، کارمند کتابخانه ملی ایران

چکیده

امروزه شبکه‌های اجتماعی به عنوان یکی از اصلی‌ترین منابع تولید داده در جهان شناخته می‌شوند. یکی از مهم‌ترین داده‌های تولید شده توسط این رسانه‌ها، کلان داده‌ها می‌باشند. تحلیل اینگونه داده‌ها نیازمند سیستم های دقیق و الگوریتم‌های هوشمند برای بازیابی اطلاعات و کشف دانش پنهان از میان آن‌ها است. با توجه به افزایش و گسترش روزافزون حجم اطلاعات منتقل شده در شبکه‌های اجتماعی، مدیریت آن‌ها به مسئله بزرگی برای سازمان‌های مختلف از جمله کتابخانه‌ها تبدیل شده است. این اطلاعات مفید برای سازمان‌ها با هدف به دست آوردن بینش غنی‌تر و عمیق‌تر برای تصمیم‌گیری‌های خرد و کلان استفاده می‌شود. به همین دلیل پیاده‌سازی، پردازش و کشف دانش از کلان داده‌های بزرگ تا جایی که ممکن است، نیاز به تحلیل و اجرای دقیق دارد. هدف این پژوهش بررسی کلان داده‌ها و تحلیل این داده‌های تولید شده در شبکه‌های اجتماعی است و سپس نقش این داده‌ها را در کتابخانه‌ها بررسی کرده است.

واژه‌های کلیدی: کلان داده، شبکه اجتماعی، تحلیل، کتابخانه

[Downloaded from jnrihs.ir on 2026-11-18]

۱- مقدمه

فناوری اطلاعات یکی از عواملی است که در همه امور زندگی اجتماعی و فردی بشر تأثیر دارد و در سال‌های اخیر مورد توجه دولت‌ها قرار گرفته است. کشورها برای جبران عقب‌ماندگی‌ها و زندگی بهتر و نیز برای کسب قدرت بیشتر در رقابت‌های بین‌المللی و حفظ تفوق خود به توسعه فناوری در حوزه‌های مختلف اقدام نموده‌اند. امروزه تحولات فناورانه تمام ابعاد سیاسی، امنیتی، اجتماعی، اقتصادی و فرهنگی را درنوردیده و به موازات آن الگوهای رفتاری و شیوه‌های ارتباطی نیز متحول شده‌اند (محمودزاده و همکاران، ۱۳۹۶).

با گسترش استفاده از اینترنت بین اقشار مختلف، استقبال کاربران از شبکه‌های اجتماعی روزبه‌روز در حال افزایش است. بر اساس آمارهای ارائه شده در ژانویه ۲۰۱۶، تعداد کل جمعیت جهان ۷. ۳۹۵ میلیارد نفر بوده است. در سال ۲۰۱۷ از این تعداد، بیش از سه میلیارد و هفتصد میلیون نفر کاربر اینترنت و از بین این کاربران ۲. ۳۰۷ میلیارد نفر از جمله کاربران فعال شبکه‌های اجتماعی[1]هستند (نویدی و همکاران، ۱۳۹۶).

از دلایل محبوبیت استفاده از این رسانه‌ها این است که کاربران فرصت دریافت یا ایجاد و به اشتراک گذاشتن پیام‌های عمومی را با هزینه‌های کم و در همه‌جا را خواهند داشت. رشد بسیار زیاد استفاده از این قبیل شبکه‌ها منجر به انباشت داده‌های عظیم شده است که آن‌ها را کلان داده[2]نامیده‌اند (استیگلیتز و همکاران[3] ۲۰۱۸).

کلان داده به مجموعه داده‌هایی گفته می‌شود که مدیریت کنترل و پردازش آن‌ها فراتر از توانایی ابزارهای نرم‌افزار در یک زمان قابل تحمل و مورد انتظار است و این اصطلاح برای رشد قابل توجه در تولید، دسترس‌پذیری، ذخیره‌سازی و پردازش داده‌ها و اطلاعات به کار می‌رود. از جمله ویژگی‌های اصلی کلان داده‌ها، حجم بالا، سرعت غیرقابل پیش‌بینی جریان داده‌ها و تنوع داده‌های در دسترس است؛ بنابراین حجمی از داده‌ها که این روزها می‌تواند ذخیره و پردازش شده و نتایج تحلیل آن مورد استفاده قرار گیرد، در گذشته قابل تصور هم نبود (مولایی، ۱۳۹۲). سیستم عامل‌های رسانه‌های اجتماعی امکانات فراوانی از فرمت‌های داده، از جمله داده‌های متنی، تصاویر، فیلم‌ها، صداها و موقعیت‌های جغرافیایی ارائه می‌دهند (استیگلیتز و همکاران، ۲۰۱۸).

با نگاهی به حوادث اخیر در سراسر جهان مانند حضور پر سروصدای داعش در خاورمیانه، حمله انتحاری به فرودگاه آتاتورک ترکیه، حمله تروریستی در فرانسه، کودتا در ترکیه و حمله تروریستی در آلمان می‌توان به حضور پررنگ شبکه‌های اجتماعی و انتشار اطلاعات این اتفاق‌ها در فضای رسانه‌ای دنیا پی برد. لذا بررسی کارکرد کلان داده‌ها در شبکه‌های اجتماعی و نحوه تجزیه و تحلیل این‌گونه داده‌ها به منظور کشف دانش از میان آن‌ها، از اهداف اصلی این پژوهش می‌باشد.

1- Social Media
2- Big Data
3- Stieglitz et al

[Downloaded from jnrihs.ir on 2026-11-18]

۲- کلان داده

تعاریف متعددی از کلان داده موجود است. موسسه گارتنر دارایی‌های اطلاعاتی با حجم زیاد، تنوع بالا یا سرعت ایجاد زیاد را که تصمیم‌گیری و یافتن درک درست از آن نیازمند پردازش‌های اطلاعاتی نوآورانه و درعین‌حال مقرون‌به‌صرفه است را داده‌های بزرگ می‌نامد (مینلی و همکاران[1] ۲۰۱۲). از سوی دیگر شرکت IBM کلان داده را داده‌هایی می‌داند که حجم آن‌ها فراتر از حدی است که بتوان با نرم‌افزارهای رایج به مدیریت، ذخیره و تحلیل آن‌ها پرداخت. مایکل فرانکلین نیز معتقد است که کلان داده، داده‌هایی هستند که مدیریت آن‌ها مستلزم هزینه فراوان است و کسب ارزش از آن‌ها نیز دشوار است (سهرابی و ایرج، ۱۳۹۴).

با توجه به این تعاریف می‌توان کلان داده را چنین تعریف کرد: کلان داده به افزایش حجم داده‌هایی که ذخیره‌سازی، پردازش و تجزیه و تحلیل آن‌ها از طریق فناوری‌های پایگاه داده سنتی دشوار است، اطلاق می‌شود (گانتز و رینسل[2] ۲۰۱۲).

این نوع داده‌ها با چند ویژگی منحصر به‌فرد شناخته می‌شوند. در برخی پژوهش‌ها این ویژگی‌ها را به سه دسته تقسیم کرده‌اند: تنوع، حجم، سرعت (عزیزی وامرزانی و خادمی، ۱۳۹۳؛ مرکز فناوری اطلاعات اینتل[3] ۲۰۱۲؛ سینگ و سینگ[4] ۲۰۱۱؛ گرهارت و همکاران[5] ۲۰۱۲). در واقع همین سه ویژگی کلیدی است که نگهداری و پردازش داده‌های بزرگ را با مشکل مواجه می‌کند. به این مدل، مدل 3V گفته می‌شود (چن و همکاران[6] ۲۰۱۴؛ چن و ژانگ[7] ۲۰۱۴؛ هاشم و همکاران[8] ۲۰۱۵).

با توجه به ماهیت کلان داده‌ها، می‌توان این ویژگی‌ها را به این صورت توضیح داد که:

تنوع، کلان داده‌ها را واقعاً حجیم می‌کند. داده‌های حجیم از منابع بسیار متنوع به دست می‌آید و معمولاً در سه نوع می‌باشد: ساخت‌یافته، نیمه ساخت‌یافته و غیر ساخت‌یافته. داده‌های ساخت‌یافته در انباره‌های داده، تگ خورده و به آسانی ذخیره می‌شود؛ اما داده‌های غیر ساخت‌یافته به صورت تصادفی بوده و تحلیل آن‌ها دشوار است. داده‌های نیمه ساخت‌یافته از فیلدهای ثابتی تشکیل نشده‌اند اما تگ‌هایی برای جداسازی عناصر داده دارند (شریفی و سمیعی، ۱۳۹۲؛ سینگ و سینگ، ۲۰۱۲). در خصوص این ویژگی ذکر این نکته حائز اهمیت است که تنوع این داده‌ها به‌گونه‌ای است که طیف وسیعی از انواع داده‌ها از جمله داده‌های متنی، عددی، صوت، تصویر، ویدئو، داده‌های برآمده از شبکه‌های اجتماعی و. . . را شامل می‌شوند (درمنده و همکاران، ۱۳۹۵).

1- Minelli et al
2- Gantz & Reinsel
3- Center, Inet IT.
4- Singh & Singh
5- Gerhardt et al
6- Chen et al
7- Chen & Zhang
8- Hashem et al

[Downloaded from jnrihs.ir on 2026-11-18]

حجم و اندازه داده‌های امروز بزرگ‌تر از ترابایت و پتابایت می‌باشد. با افزایش حجم داده، روش‌های ذخیره و فنون تحلیل سنتی داده غیرقابل استفاده می‌شوند (شریفی و سمیعی، ۱۳۹۲؛ مادن[1]، ۲۰۱۲). در حقیقت حجم داده‌ها به مقیاس قابل توجه داده‌ها اشاره دارد و منبع تولید این حجم عظیم داده‌ای متنوع شامل انسان و ماشین است (درمنده و همکاران، ۱۳۹۵).

سرعت، که نه تنها برای داده‌های حجیم بلکه برای همه فرایندها لازم است. برای فرایندهای محدود به زمان، در کلان داده‌ها، می‌بایست جریان داده به داده سازمان‌یافته تبدیل گردد تا ارزش آن به حداکثر برسد (شریفی و سمیعی، ۱۳۹۲؛ مادن، ۲۰۱۲). داده‌های بزرگ توسط حسگرهای متنوع و برنامه‌های کاربردی فراوانی به‌طور دائمی و پیوسته در حال تولید است و سرعت این تولید نیز بسیار زیاد است در نتیجه این سرعت، پردازشی سریع نیز باید برای آن‌ها وجود داشته باشد زیرا ممکن است تأخیر در پردازش داده‌ها سبب از دست رفتن فرصت و تأخیر در تصمیم‌گیری شود (درمنده و همکاران، ۱۳۹۵؛ علیزاده و همکاران، ۱۳۹۵).

با توجه به توسعه روزافزون کلان داده‌ها و به دلیل ایجاد تمایز بیشتر این داده‌ها با سایر داده‌های موجود، پژوهشگران و متخصصان کلان داده، تصمیم گرفتند تا ویژگی‌های بیشتری را به ویژگی‌های قبل اضافه کنند. یکی از این ویژگی‌ها «صحت» است که به بر مبنای اعتبار و درستی این‌گونه داده‌ها، انتخاب شده است. ویژگی دیگر «ارزش» است که به دیدگاه فناوری به ارزش از نگاه و کسب وکار اشاره دارد (گانتز و رینسل، ۲۰۰۷). ویژگی «اعتبار» که بر کاربرد درست داده‌ها تأکید دارد و «نوسان» که به این مسئله اشاره می‌کند که کلان داده‌ها برخلاف دیگر داده‌ها با گذشت زمان و انباشته شدن، ارزشمندتر می‌شوند و می‌توانند به تصمیم‌گیری‌های دقیق‌تری بینجامند (درمنده و همکاران، ۱۳۹۵؛ سان[2]، ۲۰۱۸).

بنابراین می‌توان گفت که کلان داده‌ها، مجموعه از دارایی‌های اطلاعات با مشخصه‌هایی همچون حجم، سرعت و تنوع زیاد است که نمی‌توان با روش‌های سنتی فناوری اطلاعات و ارتباطات، نرم‌افزارها و سخت‌افزارهای موجود در زمانی معقول به مدیریت و تحلیل آن‌ها پرداخت، بلکه کسب ارزش از آن‌ها نیازمند صرف هزینه و به‌کارگیری روش‌ها و فناوری‌های نوین است (درمنده و همکاران، ۱۳۹۵).

۳- فرمت کلان داده

کلان داده‌ها از نظر فرمت به سه دسته کلی تقسیم می‌شوند:

1- Madden
2- Sun

[Downloaded from jnrihs.ir on 2026-11-18]

داده ساخت‌یافته

داده‌های ساخت‌یافته، داده‌هایی با ساختار محکم و ثابت می‌باشند که می‌توان آن‌ها را در قالب جداول مدل رابطه‌ای یا اشیاء مدل شیءگرایی گنجاند. روی این‌گونه داده‌ها می‌توان پرس و جوهای بیانی مانند SQL و QQL را مطرح ساخت (ذوالفقارنسب، ۱۳۹۶).

داده نیمه ساخت‌یافته

داده‌های نیمه ساخت‌یافته، داده‌هایی دارای ساختار ضعیف هستند ولی یا بی‌قاعده هستند و یا مداوم تغییر می‌کنند که نمی‌توان این ساختار را در قالب بانک‌های رابطه‌ای یا شیءگرا بیان کرد (ذوالفقارنسب، ۱۳۹۶).

داده غیر ساخت‌یافته[1]

داده‌های غیر ساخت‌یافته، داده‌هایی می‌باشند که دارای هیچ شمای مشخصی نمی‌باشند. این داده‌ها به‌صورت متنی یا صفحات HTML می‌باشند که درصد زیادی از داده‌های موجود در وب را تشکیل می‌دهند. جستجو روی چنین داده‌هایی به‌صورت جستجو بر اساس کلمه کلیدی می‌باشد (ذوالفقارنسب، ۱۳۹۶).

٤- منابع کلان داده

منابع تولید کلان داده‌ها شامل داده‌های اینترنتی، داده‌های کسب شده از حسگرها و اطلاعات ذخیره شده از تراکنش‌ها است که کلیه داده‌های غیر ساخت‌یافته تا داده‌های ساخت‌یافته را در فرمت‌های مختلف دربرمی‌گیرد (افسری شولی و همکاران، ۱۳۹۵). در نتیجه تنوع گسترده منابع داده، داده به‌دست‌آمده از نظر افزونگی، ثبات و نویز متفاوت است. انواع مختلف منابع تولید کلان داده‌ها به شرح زیر است:

رسانه‌های اجتماعی

اطلاعاتی است که از طریق به اشتراک‌گذاری و یا تبادل اطلاعات از طریق آدرس‌های اینترنتی و یا از طریق ارتباطات مجازی و شبکه‌های مجازی به دست می‌آیند، نظیر اطلاعاتی که در پروژه‌های اشتراکی، بلاگ‌ها، میکروبلاگ‌ها، فیس‌بوک، توییتر و. . . تولید می‌شوند (هاشم و همکاران، ۲۰۱۵).

1- Unstructured data

[Downloaded from jnrihs.ir on 2026-11-18]

داده‌های ماشین

اطلاعاتی است که به صورت خودکار توسط سخت‌افزار و نرم‌افزارهای ابزارهایی نظیر کامپیوترها، وسایل پزشکی یا دیگر ماشین‌ها بدون دخالت انسان تولید می‌شود (هاشم و همکاران، ۲۰۱۵).

حسگرها

وسایل حسگر متفاوتی برای اندازه‌گیری کمیت‌های فیزیکی و تبدیل آن‌ها به سیگنال وجود دارد که بخشی از داده‌های عظیم را تولید می‌نماید (هاشم و همکاران، ۲۰۱۵).

تراکنش‌های اینترنتی (IoT)

IoT یک مجموعه از اشیایی است که به صورت یکتا قابل تعریف هستند و به عنوان بخشی از اینترنت می‌باشند. این اشیا شامل تلفن‌های کوچک، دوربین‌های دیجیتال و تبلت‌ها هستند. وقتی این وسایل از طریق اینترنت به یکدیگر متصل می‌شوند، قادرند بیشتر پردازش‌های کوچک و سرویس‌های پشتیبانی پایه اقتصادی، محیطی و سلامت مورد نیاز را فراهم آورند. تعداد زیاد وسایل متصل به اینترنت، انواع مختلفی از سرویس‌ها را فراهم می‌آورند و مقادیر زیادی داده و اطلاعات تولید می‌نمایند (رائو و همکاران[1]، ۲۰۱۲).

۵- پردازش کلان داده‌ها

کلان داده‌ها می‌توانند بر اساس سه بُعد حجم، تنوع و نرخ تولید تعریف شوند. اندازه داده‌ها، ارزش و پتانسیل داده‌های مورد توجه به آن را تعیین می‌کند تا جایی که می‌توان تصمیم گرفت که کلان داده محسوب شود یا خیر. جنبه بعدی در داده‌های حجیم، تنوع آن‌هاست. دسته‌بندی کردن کلان داده‌ها نیاز به شناسایی توسط تحلیلگران داده دارد. نرخ تولید به چگونگی سرعت تولید و پردازش داده‌ها برای پاسخگویی به خواسته و چالش‌های پیش رو در مسیر رشد و توسعه اشاره دارد.

داده‌ها در این سه بُعد در حال رشد هستند و روزبه‌روز نرخ رشد آن‌ها با سرعتی باور نکردنی در افزایش است و سه ویژگی حجیم بودن، متغیر بودن و بدون ساختار بودن کلان داده‌ها باعث شده است که تکنولوژی نوین مدیریت پایگاه داده‌ها در قالب پروژه متن‌باز هادوپ[2] ارائه می‌شود. هادوپ برای تحلیل داده‌ها در یاهو شکل گرفت و سپس آن را به آپاچی[3] بخشید تا به رایگان در دسترس عموم قرار گیرد و توسعه‌دهندگان مختلفی در سراسر دنیا توانایی‌های آن را روزبه‌روز بهتر کنند. رمز موفقیت هادوپ، شکستن ساختار داده به بخش‌های کوچک‌تر و پردازش هر بخش به صورت جداگانه روی سرورهای ابری است. این کار تنها هزینه نگهداری و پردازش را به شدت کاهش می‌دهد، بلکه با تقسیم داده‌ها در بخش‌های کوچک‌تر، پردازش داده‌هایی که تا پیش از این غیرممکن بود را محقق می‌سازد.

1- Rao et al
2- Hadoop
3- Apache

[Downloaded from jnrihs.ir on 2026-11-18]

کلان داده‌ها به فناوری‌های پیشرفته‌ای برای پردازش شدن، به روشی کارآمد و در زمانی معقول نیاز دارند. فناوری‌هایی که به‌طور عمومی در این زمینه استفاده می‌شوند، شامل مواردی مانند پایگاه‌های داده با قابلیت پردازش موازی عظیم، دسته‌بندی‌های داده‌کاوی، سیستم‌های فایلی توزیع شده، پلتفرم‌های پردازش ابری، اینترنت و همچنین سیستم‌های ذخیره‌سازی مقیاس‌پذیر هستند. (افسری شولی و همکاران، ۱۳۹۵).

۱-۵ آپاچی هادوپ

هادوپ یک سکوی نرم‌افزار متن‌باز می‌باشد که از طریق آن می‌توان به سادگی حجم بسیار بزرگی از داده‌ها را ذخیره، تحلیل و پردازش کرد. برای مثال هادوپ می‌تواند برای کارهایی نظیر شاخص‌گذاری صفحات وب، داده‌کاوی، تحلیل فایل‌ها، یادگیری ماشین، تحلیل‌های مالی، شبیه‌سازی‌های علمی و یا تحقیقات در زمینه بیوانفورماتیک مورد استفاده قرار بگیرد (افسری شولی و همکاران، ۱۳۹۵).

در چهار نوع ویژگی می‌توان هادوپ را بررسی کرد که در بحث مقیاس‌پذیری می‌توان گفت وسط سکوی هادوپ این امکان وجود دارد که مجموعه داده‌هایی با حجمی در مقیاس چندین پتابایت داده را توسط توزیع داده‌ها در مقیاس چندین هزار گره در یک خوشه را مدیریت و پردازش کرد. در بحث کارآمدی می‌توان گفت پردازش موازی داده‌ها و وجود یک سیستم فایل توزیع شده این امکان را فراهم می‌کند که بتوان با سرعت بیشتری داده‌ها را پردازش کرد. در حوزه قابلیت اعتماد بودن می‌توان این جمله را آورد که: چندین تکرار از داده‌ها می‌تواند ایجاد و در سطح خوشه توزیع شود. در هنگام ایجاد خطا در یکی از گره‌ها، پلتفرم می‌تواند بدون دخالت کاربر به سازمان‌دهی مجدد بپردازد. در نتیجه تصحیح خطا خودکار نیز امکان‌پذیر است. در بحث سادگی هم این را می‌توان گفت که با هادوپ، نوشتن کدهای موازی کارا به سرعت و سهولت میسر است و جزئیات پردازش موازی، مانند توزیع داده میان گره‌های پردازشی، راه‌اندازی مجدد زیر وظیفه‌های از کار افتاده و ذخیره نتایج نهایی محاسبات در سیستم فایل، از دید کاربر پنهان می‌ماند. همچنین حتی دانشجوها هم می‌توانند به سرعت و با هزینه پایین برای خود یک خوشه هادوپ ایجاد کنند (یوسفی کندول، ۱۳۹۷).

از بُعد معماری می‌توان گفت که معماری هادوپ یک معماری Master/Slave می‌باشد. در این معماری، بر روی ماشینی که به عنوان Master انتخاب می‌شود، سرویس‌هایی برای مدیریت و پیکربندی کل سیستم و مدیریت برنامه‌ها اجرا می‌شوند. همچنین در ماشین Master ابزارهایی برای مدیریت داده‌ها و تحویل برنامه به سیستم وجود دارد که کاربران سیستم برای کار با هادوپ می‌توانند از آن‌ها استفاده می‌کنند. در ماشین Slave نیز سرویس‌هایی برای اجرای وظایفی که از سمت ماشین Master به آن ارسال می‌شود، به اجرا درمی‌آیند (افسری شولی و همکاران، ۱۳۹۵).

سکوی هادوپ از دو قسمت اصلی تشکیل شده است که این قسمت‌ها شامل سیستم فایل توزیع شده هادوپ (HDFS) و چارچوب برنامه‌نویسی Map Reduce می‌باشند. HDFS وظیفه ذخیره‌سازی و

[Downloaded from jnrihs.ir on 2026-11-18]

مدیریت داده‌ها در سیستم فایل را برعهده داشته و داده‌های ورودی را به بلوک‌های مساوی تقسیم می‌کند و به ماشین‌های Slave تحویل می‌دهد. این سیستم فایل از طریق تکرار داده‌ها به مقابله با خطا می‌پردازد. چارچوب Map Reduce نیز یک مدل برنامه‌نویسی با هدف پردازش و تولید مجموعه عظیمی از داده‌ها فراهم می‌کند (یوسفی کندول، ۱۳۹۷).

سیستم فایل توزیع شده هادوپ و چارچوب برنامه‌نویسی Map Reduce که اجزای اصلی هادوپ را تشکیل می‌دهند، مبین این هستند که هادوپ سیستمی است برای موازی‌سازی کارها و کارایی حقیقی آن در محیطی توزیع شده مشخص می‌گردد. با این حال در کارهای تست و توسعه همیشه فراهم آوردن چنین زیرساختی آسان نیست. از این رو است که هادوپ به چند شکل قابل استقرار است. سه حالت نصب و راه‌اندازی برای هادوپ در نظر گرفته شده است. در حالت تک گره یا عملکرد مستقل شما می‌تواند روی یک گره (مثلاً یک لپ‌تاپ) هادوپ را نصب کنید. در حالت شبه توزیع شده نیز این امر میسر است با این تفاوت که هر یک از پردازه‌های هادوپ روی یک پردازه جاوای مستقل اجرا خواهند شد. حالت سوم هم وضعیت تمام توزیع شده آن می‌باشد که نصب در این حالت بر روی خوشه خواهد بود (افسری شولی و همکاران، ۱۳۹۵).

٦- کاربردهای کلان داده

سالانه هزینه‌های هنگفتی صرف جمع‌آوری و بایگانی داده‌های مختلف می‌شود اما در عمل این حجم عظیم از داده چه کاربردی دارد و چرا امروزه روش‌های تجزیه و تحلیل کلان داده‌ها توجه زیادی را به خود معطوف کرده‌اند. در این قسمت به نمونه‌هایی از کاربرد این داده‌ها اشاره شده است:

سیاست و حکومت

تحلیل کلان داده‌ها نقش بسیار مهمی در هدایت استراتژی مبارزات انتخاباتی باراک اوباما در سال ۲۰۱۲ در ایالات متحده آمریکا داشت. به‌گونه‌ای که وی یک تیم تحلیلی کلان داده‌های رسانه‌های اجتماعی داشت (از جمله فیس‌بوک و توییتر) که می‌توانستند ویژگی‌ها و رفتار رأی‌دهندگان را تحلیل کنند و تأثیر بسزایی بر روند انتخابات داشتند (کوماری[1] ۲۰۱۶).

بازاریابی

با توجه به این‌که حجم کلان داده‌ها روزبه‌روز بیشتر خواهد شد، این محتوا سبب جاری شدن سیل بی‌پایان ثروت دیجیتال به سمت فروشگاه‌هایی خواهد شد که از این داده‌ها به‌درستی استفاده کنند. اطلاعات سنتی که از سوی شرکت‌ها، صنایع، مؤسسات و سازمان‌های غیرانتفاعی تأمین می‌شود. به این اطلاعات، داده‌های بسیار زیاد شبکه‌های اجتماعی که از طریق وبلاگ‌ها، گروه‌های بحث، ویدئوها، صوت و تصاویر نیز

1- Kumari

[Downloaded from jnrihs.ir on 2026-11-18]

افزوده می‌شود. این ترکیب اطلاعات سنتی به‌علاوه این داده‌ها، می‌تواند بر منافع، چشم‌اندازها و بینش در مورد نحوه کسب‌وکار، مصرف‌کنندگان و سازمان‌های ارائه‌کننده خدمات و محصولات و حتی اقتصاد جهانی تأثیرگذار باشد. به نحوی‌که در برابر کاهش هزینه‌ها، بهره‌وری سازمان‌ها و شرکت‌های خدماتی افزایش پیدا کند و از سوی مدیران تصمیمات بهتری گرفته شود. از سویی کلان داده‌ها در حوزه رقابتی، تهدید بزرگی برای کسب‌وکارهایی است که آن را نادیده می‌گیرند. بسیاری از مدیران رده بالا و بازاریابان به شرکت‌ها و سازمان‌ها کمک می‌کنند تا چگونه از کلان داده‌ها استفاده کنند تا آن را به عنصر مهمی از مجموعه مهارت‌های حرفه‌ای و بازاریابی تبدیل کنند (کوماری، ۲۰۱۶).

آموزش و پرورش

حجم بسیار زیادی از داده‌ها در مدارس و مؤسسات آموزش عالی تولید می‌شود. این داده‌ها سبب درک الگوهای عملکردی و رفتاری دانش آموزان می‌شود. همچنین به دانش آموزان اجازه می‌دهد محتوای شخصی و روش‌های آموزشی مختص به خود را داشته باشند. معلمان نیز می‌توانند پاسخ‌ها به سؤالات امتحان را تحلیل کرده و به آن‌ها کمک کنند تا در مورد مفاهیم خاص به درک درستی برسند. همچنین تحلیل مفاهیم درسی در کتاب‌ها یا ابزارهای یادگیری آنلاین سبب آشنایی با نحوه کار آن‌ها و تجدیدنظر در آن‌ها خواهد شد. در نهایت می‌تواند سبب تعامل هوشمند شود به‌گونه‌ای که اثربخشی و فنون مختلف یادگیری را افزایش یابد (کوماری، ۲۰۱۶؛ دهوانان و زنینی[1] ۲۰۱۴).

۷- چالش‌های کلان داده

بسیاری از پژوهشگران بُعد ابعاد پذیری را به عنوان یک جنبه از مشکلات کلان داده‌ها در نظر می‌گیرند. در واقع کلان داده‌ها نباید در حجم داده‌های معمولی فشرده شوند، بلکه همه ویژگی ابعاد بزرگ داده‌ها باید در نظر گرفته شود. در واقع پردازش داده‌ها با ابعاد بالا در حال حاضر یک امر دشوار در پژوهش‌های علمی محسوب می‌گردد. برای مقابله با این چالش، تکنیک‌های پیشرفته‌ای به وجود آمده‌اند تا نقشه فضای داده با ابعاد بالا را با حداقل فقدان اطلاعات به فضایی با ابعاد کمتر تبدیل شود. یکی از محبوب‌ترین این روش‌ها، روش تحلیل مؤلفه‌های اصلی است (زاهد و سخی، ۱۳۹۵).

یکی دیگر از چالش‌های موجود در کلان داده‌ها، عدم تعامل سیستم رایانه با سرعت تولید داده‌ها است. افزایش فزایند حجم داده‌ها، مشکلات زیادی در جمع‌آوری، ذخیره‌سازی و مدیریت و حفظ امنیت داده‌ها ایجاد می‌کند. همچنین نرخ رشد اطلاعات به صورت نمایی است، درحالی‌که شیوه‌های پردازش اطلاعات به کندی پیشرفت می‌کنند (سنمی علمداری، ۱۳۹۵).

امروزه به منظور حل چالش ذخیره‌سازی حجم عظیمی از اطلاعات از رایانش ابری استفاده می‌شود. رایانش ابری مدلی است برای فراهم کردن دسترسی آسان بر اساس تقاضای کاربر از طریق شبکه به مجموعه‌ای از

1- Dhawan & Zanini

[Downloaded from jnrihs.ir on 2026-11-18]

منابع محاسباتی قابل تغییر و پیکربندی مانند شبکه‌ها، سرورها، منابع ذخیره‌سازی، برنامه‌های کاربردی و خدماتی که این دسترسی بتواند با کمترین نیاز به مدیریت منابع و یا نیاز به دخالت مستقیم ارائه‌دهنده خدمات به سرعت فراهم‌شده یا عرضه گردد (مل و گرنس[1] ۲۰۰۹).

اگرچه استفاده از رایانش ابری و ذخیره‌سازی اطلاعات در ابر تا حدی مشکل دسترسی به داده‌ها و ذخیره‌سازی اطلاعات را حل کرده است، اما هنوز امنیت اطلاعات در رایانش ابری به عنوان یک چالش مهم تلقی می‌شود تا جایی که پژوهش جدیدی توسط مایکروسافت نشان داد که ۵۱ درصد از صاحبان کسب وکارها به خاطر نگرانی از امنیت داده‌ها نسبت به استفاده از راهکارهای مبتنی بر رایانش ابری بی‌میل هستند (میرعابدینی و روشن ضمیر، ۱۳۹۵)؛ اما به‌رغم این ترس‌ها واقعیت چیز دیگری است: رایانش ابری می‌تواند امنیت بسیار بیشتری را نسبت به آنچه خود سازمان‌ها می‌توانند انجام دهند برای داده‌های آن‌ها تأمین کند.

به جز امنیت به عنوان مهم‌ترین چالش رایانش ابری، این تکنولوژی با دو چالش دیگر نیز روبرو است: نیاز به اتصال دائمی به اینترنت و توافق بر سطح خدمات. درصورتی‌که کاربر نتواند به اینترنت متصل شود، رایانش ابری غیرممکن خواهد بود. ازآنجایی‌که کاربر باید برای ارتباط با برنامه‌های کاربردی و اسناد خود به اینترنت متصل باشد. اگر ارتباط اینترنتی نداشته باشد نمی‌تواند به هیچ چیزی حتی اسناد خود دسترسی پیدا کند (بهشتی و سروری، ۱۳۹۱). همچنین کاربران استفاده از رایانش ابری هیچ کنترلی روی زیرساخت‌های محاسباتی ندارند، از این رو بایستی یکسری تضمین در خصوص عوامل کیفیتی، قابلیت اطمینان، دسترس‌پذیری و کارایی منابع به مشتری (علی‌الخصوص مشتریانی که کسب وکارهای خود را توسط این خدمات انجام می‌دهند) داده شود (مقدس و آسا، ۱۳۹۲).

۸- شبکه‌های اجتماعی

شبکه‌های اجتماعی مجازی، نسل جدیدی از وب‌سایت‌های اینترنتی هستند. در این وب‌سایت‌ها، کاربران اینترنتی حول محور مشترکی به صورت مجازی دور یکدیگر جمع می‌شوند و جماعت‌های آنلاین را تشکیل می‌دهند (مولایی، ۱۳۹۲). در واقع شبکه اجتماعی شبکه‌ای متشکل از افراد و گروه‌ها و ارتباطات بین آن‌هاست (نویدی و همکاران، ۱۳۹۶).

یک شبکه اجتماعی، سایت یا مجموعه سایتی است که به کاربران که علاقه به اشتراک‌گذاری علاقه‌مندی‌ها، افکار و فعالیت‌های خود با دیگران دارند، این امکان را می‌دهد. این‌گونه رسانه‌ها طی دهه گذشته تبدیل به یک ابزار مهم جهت دستیابی و نشر اطلاعات در حوزه‌های مختلف مانند کسب‌وکار، سرگرمی، مدیریت بحران و سیاست تبدیل شده‌اند (استیگلیتز و همکاران، ۲۰۱۸). شبکه‌های اجتماعی در عین حالی که فضایی برای پیدا کردن دوستان جدید هستند، به افراد این امکان را می‌دهند که با دوستان

1- Mell & Grance

[Downloaded from jnrihs.ir on 2026-11-18]

قدیمی خود ارتباط برقرار کنند. همچنین مکان‌هایی برای تبادل‌نظر هستند که در آن‌ها افراد عقاید و نظرات خود را با هم به اشتراک می‌گذارند. به‌طورکلی شبکه‌های اجتماعی مکان‌هایی در دنیای مجازی هستند که مردم خود را به‌طور خلاصه معرفی می‌کنند و امکان برقراری ارتباط بین خود و هم‌فکرانشان را در زمینه‌های مختلف موردعلاقه فراهم می‌کنند. در واقع شکل جدیدی از برقراری ارتباط و به اشتراک‌گذاری محتوا در اینترنت است. شبکه‌های اجتماعی پایگاه‌های جدیدی هستند که در مرکز توجه کاربران اینترنت قرار گرفته‌اند و هرکدام دسته‌ای از کاربران با ویژگی خاص را گرد هم می‌آورند (اسلامی، ۱۳۹۱).

زمان زیادی از فراگیر شدن شبکه‌های اجتماعی و گستردگی دامنه استفاده از آن نمی‌گذرد اما سرعت جذب کاربران آن از برآوردهای اولیه، بسیار بیشتر بوده است. این شبکه‌ها با قدمت کمی که دارند، توانسته‌اند بر بسیاری از جوانب زندگی بشر از جمله زندگی فردی، اجتماعی، اقتصادی، سیاسی و فرهنگی تأثیرگذار باشند و با نادیده گرفتن مرزهای سیاسی و جغرافیایی، زمینه‌های همگرایی و یکپارچگی جهانی را فراهم سازند. ازآنجایی‌که یکی از بسترهای مهم جریان داده‌ها، شبکه‌های اجتماعی است دستیابی به این داده‌ها و نحوه تحلیل آن‌ها از اهمیت بسیاری برخوردار است (محمودزاده و همکاران، ۱۳۹۶). یکی از مهم‌ترین نوع داده‌های تولید شده توسط این رسانه‌ها، کلان داده‌ها هستند که به منظور تحلیل این داده‌ها لازم است از روش‌های تحلیلی مناسب استفاده کرد.

۹- تحلیل کلان داده‌های استخراجی از شبکه‌های اجتماعی

روش‌های تحلیل داده‌های استخراج شده از شبکه‌های اجتماعی را می‌توان در دو دسته خودکار و روش‌های انسانی قرار داد.

روش‌های خودکار، روش‌های نرم‌افزاری کمی هستند که بر استخراج ماشینی ساختار و واژگان شبکه تمرکز یافته‌اند و شامل روش‌های تحلیل شبکه اجتماعی، تحلیل معنایی و تحلیل تمایلات می‌شوند.

روش‌های انسانی نیز اغلب روش‌های کیفی و تفسیری هستند که مبتنی بر محتوا، تجارب کاربران و مردم‌نگاری، دیجیتال می‌شوند که توسط متخصص انسانی صورت می‌گیرند.

۹-۱- روش‌های تحلیل خودکار

این روش‌ها به بررسی روابط اجتماعی از نظر تئوری شبکه و بر اساس گره‌ها و روابط می‌پردازند و بر روابط بین موجودیت‌های اجتماعی تمرکز یافته‌اند و با استفاده از گراف‌ها و ماتریس‌ها، روابط اجتماعی شبکه را نمایش می‌دهند. نقشه‌ها و گراف‌های شبکه‌های اجتماعی، روش سریعی را برای ارزیابی انواع ساختارهای اجتماعی که حول یک موضوع پدید می‌آیند، ارائه می‌کنند (کمپل و همکاران[1]، ۲۰۱۳).

1- Campbell et al

[Downloaded from jnrihs.ir on 2026-11-18]

روش تحلیل شبکه اجتماعی

شناخته شده‌ترین روش تحلیل خودکار، روش تحلیل شبکه اجتماعی معروف به روش **SNA** است که به عنوان یک فیلد مطالعه به‌طور گسترده تحت تأثیر کارهای دانشمندان علوم اجتماعی مانند فریمن[۱](۱۹۶۳)، لومان و گاتمن[۲](۱۹۶۶) و گرانووتر[۳](۱۹۷۵) می‌باشد؛ بنابراین تحلیل شبکه اجتماعی مبتنی بر تئوری شبکه، به عنوان یک تکنیک کلیدی در جامعه‌شناسی مدرن ظهور کرده و طرفداران قابل توجهی نیز در حوزه‌های انسان‌شناسی، زیست‌شناسی، مطالعات ارتباطات، اقتصاد، جغرافیا و. . . پیدا کرده و تبدیل به یک موضوع مهم مطالعاتی شده است. بر این اساس می‌توان گفت، مطالعه شبکه‌های اجتماعی از ابتدا با تأکید بر ساختارهای اجتماعی تعریف شده است. برخی از نرم‌افزارهای توسعه یافته در این مورد پاژک[۴]، نود ایکس ال[۵]، کامتریکس[۶]، اگونت[۷] و نت ماینر[۸] می‌باشند (نویدی و همکاران، ۱۳۹۶).

روش تحلیل احساسات

تا پیش از گسترش شبکه‌های اجتماعی آنلاین، گردآوری داده‌های مربوط به عقاید افراد پرهزینه بود و معمولاً در مقیاس بسیار کوچکی به دست می‌آمد. ولی هنگامی که کاربران در اول شروع به برقراری ارتباط در سطح گسترده‌ای از طریق این کانال نمودند، شبکه‌های اجتماعی پر از داده‌هایی از رأی‌ها و اظهارات افراد شد. از این جنبه، شبکه‌های اجتماعی امکانات جدیدی را برای تعامل و اشتراک‌گذاری بلادرنگ نظرات و عقاید برای افراد ایجاد کردند و از طرف دیگر، حجم داده‌های غنی و فراوان خلق شده توسط کاربران فرصت‌های بی‌پایانی را برای مطالعه رفتارهای انسانی و اثرات این رفتارها در عرصه‌های مختلف زندگی انسانی فراهم ساخت. با توجه به اهمیت نقش تمایلات در تحلیل شبکه‌های اجتماعی، در سال‌های اخیر روش‌های نرم‌افزاری مبتنی بر تحلیل احساسات متن توسعه یافته‌اند. به‌طورکلی اساس کار این روش‌ها، تحلیل احساسات بر اساس فرهنگ لغات، شمارش بسامد کلمات و روش‌های آماری است. برخی از این نرم‌افزارها شامل سنتی استرنگت[۹]، سنتی منت ۱۴۰[۱۰]، آپینیون فایندر[۱۱]، سنتی وردنت[۱۲]، سنتیک نت[۱۳] می‌شوند (نویدی و همکاران، ۱۳۹۶).

1- Freeman
2- Luman & Gotman
3- Granovetter
4- Pajek
5- NodeXL
6- Commetrix
7- EgoNet
8- NetMiner
9- Sentistrength
10- Sentiment 140
11- OpinionFinder
12- SentiWordNet
13- SenticNet

[Downloaded from jnrihs.ir on 2026-11-18]

روش تحلیل معنایی

با توجه به نتایج ضعیف تکنیک‌های لغوی که بر اساس شمارش بسامد کلمات و روش‌های آماری است، در سال‌های اخیر روش تحلیل معنایی با رویکرد مفهوم محور به تحلیل متن با استفاده از پایگاه دانش معنایی مانند هستی‌شناسی و شبکه‌های معنایی وب می‌پردازد. هدف از این روش، تولید دانش مربوط به اشخاص و جوامع در یک شبکه اجتماعی است. این امر بر اساس معناهای موجود در یک رسانه که توسط بازیگران شبکه تبادل شده است، می‌باشد. روش تحلیل معنایی شبکه اجتماعی به مطالعه ابعاد مفهومی گراف‌های اجتماعی می‌پردازد؛ بنابراین اساس کار این روش، تحلیل شبکه‌های اجتماعی بر اساس چارچوب مفهومی وب معنایی و مبتنی بر تشخیص اجتماعات معنایی است (نویدی و همکاران، ۱۳۹۶).

۲-۹- روش‌های انسانی

دومین دسته از روش‌های تحلیل شبکه‌های اجتماعی، روش‌های انسان‌محور هستند که به صورت کیفی و تفسیری به تحلیل و بررسی محتوا و فرایندهای شبکه‌های اجتماعی می‌پردازند. روش تحلیل محتوا، روش تجارب کاربران و روش مردم‌نگاری دیجیتال از جمله روش‌های تحلیل انسانی به شمار می‌آیند.

روش تحلیل محتوا

تحلیل محتوا به هر رویه‌ای که برای تحلیل، تلخیص، طبقه‌بندی و استنباط کردن خصوصیاتی خاص از متن منجر می‌شود و در مواردی معناهای پنهان واقع در آن هویدا شده و گاهی امکان مقایسه چندین متن را می‌دهد، اطلاق می‌گردد (مؤمنی راد، ۱۳۹۲).

روش تحلیل محتوا، یک روش کیفی با تمرکز بر محتوای شبکه است که به استخراج معانی از محتوای ایجاد شده توسط کاربران در بافت مربوطه می‌پردازد. هدف از این روش، تفسیر داده‌های محتوایی شبکه‌های اجتماعی می‌باشد. اساس این روش بر این اصل شکل گرفته است که تمرکز صرف بر محتوای ظاهری پیام‌های موجود در شبکه چندان ارزشمند نیست و یک تحلیل‌گر باید با به‌کارگیری استنباط و قضاوت خود بتواند به توصیف و تفسیر محتوای پنهان یک پیام ارتباطی بپردازد؛ بنابراین تحلیل محتوا سعی در نتیجه‌گیری از داده‌های متنی در بافت مربوطه دارد؛ یعنی بررسی «چه کسی، چه چیزی را به چه کسی، چرا، تا چه حد و با چه تأثیری می‌گوید» (نویدی و همکاران، ۱۳۹۶).

روش تجارب کاربران

این روش یک روش کیفی و انسان‌محور است که به بررسی رفتار کاربران، نگرش‌ها و انگیزش‌های آن‌ها در تعامل با شبکه‌های اجتماعی می‌پردازد. تجارب کاربران و نگرش آن‌ها هنگامی که از یک پلتفرم استفاده می‌کنند، دیدگاه‌هایی را درباره تجارب آن‌ها ارائه می‌کند که با استفاده از مصاحبه و مشاهده رفتار کاربران می‌توان به جمع‌آوری برداشت‌های کیفی کاربران پرداخت. بررسی‌ها نشان می‌دهد که مصاحبه‌ها

[Downloaded from jnrihs.ir on 2026-11-18]

برای تحلیل انتشار ایده‌ها و انگیزه اشتراک‌گذاری دانش مؤثر است ولی گردآوری این تجارب نیازمند یک تلاش بزرگ و منابع بسیار زیادی است. این بررسی از طریق مشاهده رفتار کاربران و اجرای مصاحبه‌های نیمه ساختاریافته برای بررسی فعالیت‌ها، جایگاه و تجارب کاربران در تعامل با شبکه‌های اجتماعی صورت می‌گیرد و نیازمند قضاوت‌های تحلیلگر انسانی است (نویدی و همکاران، ۱۳۹۶).

روش مردم‌نگاری دیجیتال

این روش، مردم‌نگاری نوین دیجیتال است. این روش یکی از روش‌های پژوهش کیفی است که با گردآوری داده‌ها از طریق ارتباطات با واسط کامپیوتری یا فناوری‌های دیجیتال به بررسی فرهنگ جوامع و سازمان‌های مجازی می‌پردازد. هدف از این روش، بررسی و تحلیل کیفی شبکه‌های انسانی از ابعاد مختلف و توجه به ارتباطات و فرایندهای روزمره تعامل و مباحثات انسانی به منظور درک چگونگی ایجاد هویت جمعی و معناسازی مشترک در این شبکه‌ها و پیوند یافتن این شبکه‌ها با اقدامات و فعالیت‌های این افراد می‌باشد (نویدی و همکاران، ۱۳۹۶).

۱۰- کاربرد نتایج تحلیل شبکه‌های اجتماعی

حوزه‌های کلیدی کاربرد روش‌های تحلیل شبکه اجتماعی مبتنی بر تئوری شبکه، شامل جریان تبادل منابع، تبادل بین کشورها، جریان اطلاعات و ایده‌ها، اشاعه نوآوری در سازمان‌ها، شیوع و نفوذ بیماری‌ها و جریان حمایت اجتماعی است.

نتایج حاصل از تحلیل احساسات و تحلیل معنایی نیز می‌تواند در بازاریابی محصولات و خدمات، پیش‌بینی گرایش‌های سیاسی و نتایج انتخابات مفید واقع شود؛ ولی حوزه تحلیل کیفی شبکه کمتر علاقه‌مند به تبادل منابع و بیشتر متمرکز بر بررسی تجارب زیستی شبکه‌های اجتماعی، آنچه در شبکه‌ها می‌گذرد، تعاملات، فرایندها و عواقب ناشی از پویایی شبکه برای نابرابری در زندگی اجتماعی است. روش‌های کیفی برای شش جنبه از تحلیل شبکه‌های اجتماعی مناسب هستند: شناسایی شبکه‌ها، بررسی فعالیت‌های شبکه، تفسیر شبکه، اثرات شبکه، پویایی شبکه و دسترسی به بازیگران و شبکه‌ها (نویدی و همکاران، ۱۳۹۶).

۱۱- پیشینه مرتبط با تحلیل کلان داده‌های شبکه‌های اجتماعی

در خصوص تحلیل کلان داده‌ها و داده‌های شبکه‌های اجتماعی پژوهش‌های متعددی انجام شده است. موزنر و همکاران[1] در سال ۲۰۱۸ پژوهشی را با هدف تجزیه و تحلیل متن و شبکه یک انجمن اختلال خوردن که از بین ۲۴۴۷ پست و ۳۴۱۱۸ توسط ۳۰۲۹ کاربر در انجمن Reddit را انجام دادند و نتایج آن حاکی از استخراج ۹ موضوع مرتبط با حمایت اجتماعی و محتوای اختلالات خوردن بود. تحلیل شبکه‌های

1- Moessner et al

[Downloaded from jnrihs.ir on 2026-11-18]

اجتماعی نیز الگوهای میان ارتباطات و کاربران بانفوذ در این شبکه را مشخص نمود. کیم و همکاران[1]در سال ۲۰۱۷، مطالعه خود را با هدف بررسی انتقادی تحقیقات در حوزه سوءمصرف مواد و اعتیاد با استفاده از داده‌های بزرگ رسانه‌های اجتماعی انجام دادند. داده‌های جمع‌آوری‌شده از طریق پایگاه پابمد، کل مجموعه الکترونیکی "Infodemiology and Infoleillance" در مجله تحقیقات پزشکی JMIR Publications بود. خصوصیات کاربر، ویژگی‌های ارتباطی، مکانیسم‌ها و پیش‌بینی کننده‌های چنین ارتباطی و پیامدهای روانی و رفتاری رسانه‌های اجتماعی برای ارتباطات مربوط به استفاده از مواد مخدر مشکل‌ساز در این داده‌های موردبررسی قرار گرفت. نتایج این پژوهش حاکی از این مطلب بود که داده‌های بزرگ رسانه‌های اجتماعی می‌توانند یک منبع عظیم برای درک، نظارت و دخالت در سوءمصرف مواد و مشکلات اعتیاد باشند. پژوهش دیگری در سال ۲۰۱۷ انجام گرفت که در آن تحلیل شبکه‌های اجتماعی را با استفاده از کلان داده‌ها و پلتفرم InfoSphere BigInsights و Apache Flume صورت گرفت. داده‌های جمع‌آوری‌شده این پژوهش از بزرگ‌ترین وب‌سایت شبکه اجتماعی، توییتر، بود. نتایج پردازش اطلاعات جمع‌آوری‌شده از طریق Apache Flume با استفاده از InfoSphere BigInsights و ذخیره‌سازی در Hadoop نشان داد که در این پلتفرم جدید تحلیل کلان داده‌های شبکه‌های اجتماعی بهتر از سایر روش‌ها انجام می‌شود (بیرجالی و همکاران[2]، ۲۰۱۷). نتایج پژوهشی در حوزه صنعت نفت و گاز نشان داد که تجزیه‌وتحلیل داده‌های لرزه‌ای و داده‌های لرزه‌خیز، بهبود ویژگی‌ها و شبیه‌سازی مخزن، کاهش زمان حفاری و افزایش ایمنی حفاری، بهینه‌سازی عملکرد پمپ‌های تولید، بهبود مدیریت دارایی‌های پتروشیمی، بهبود حمل‌ونقل و بهبود ایمنی شغلی از مواردی هستند که با استفاده از تحلیل کلان داده‌ها به دست می‌آیند اما با این حال کیفیت داده‌ها و درک پیچیدگی آن‌ها از موضوعات چالش‌برانگیز کلان داده‌ها در این حوزه هستند (محمدپور و ترابی[3]، ۲۰۱۸).

همچنین در حوزه پزشکی از تحلیل کلان داده‌ها به منظور صحت پرتودرمانی سرطان استفاده شده است که نتایج تحلیل آن‌ها حاکی از این بود با پیشرفت فن‌آوری‌ها در بهبود کارایی، دقت و وسعت اندازه‌گیری وضعیت بیمار امکان دقت در پرتودرمانی‌ها ممکن است بهتر صورت گیرد (مک نات و همکاران[4]، ۲۰۱۸). و نیز در حوزه صنایع غذایی و مدیریت زنجیره تأمین تحلیل داده‌های شبکه اجتماعی توییتر با استفاده از روش ماشین بردار پشتیبان و خوشه‌بندی سلسله مراتبی صورت گرفت که نتایج حاوی کلماتی بود که تصمیم‌گیرندگان زنجیره تأمین در هنگام بازخورد با مشتری، مسائل مربوط به جریان و کیفیت محصولات غذایی را ارائه می‌کنند (سینگ و همکاران[5]، ۲۰۱۸).

1- Kim et al
2- Birjali et al
3- Mohammadpoor & Torabi
4- McNutt et al
5- Singh et al

[Downloaded from jnrihs.ir on 2026-11-18]

مطالعه‌ای در سال ۱۳۹۶ توسط حسینی و قلمکاری با هدف تحلیل جوامع برند آنلاین بر پایه‌ی شبکه‌های اجتماعی با روش شبکه نگاری صورت گرفت. در این پژوهش که بر پایه‌ی روش تحقیق کیفی با عنوان شبکه نگاری و با بهره‌گیری از نرم‌افزار تحلیل داده مکس کیودا انجام شده است، دو نمونه از جوامع برند در بستر شبکه‌ی اجتماعی اینستاگرام را مورد مطالعه قرار داده شد. نتایج پژوهش نشان داد، بهره‌گیری از قابلیت‌های شبکه‌های اجتماعی، منجر به نگرش مثبت به برند و ایجاد قصد خرید در مشتریان می‌شود و نظرات کاربران به نسبت محتوای تولید شده توسط بازاریابان، تأثیر بیشتری بر روی تصمیم خرید سایر اعضای صفحه برند دارد. رضوانی و کبیری (۱۳۹۶) اثرگذاری محتوای موجود در شبکه‌های اجتماعی یوتیوب و اینستاگرام بر رفتار ترافیکی شهروندان را مورد بررسی قرار دادند که نتایج این مطالعه نشان داد محتوای انگیزشی نسبت به محتوای آموزشی می‌تواند اثرگذاری بیشتری داشته باشد. همچنین محتوای ویدیویی نسبت به عکس و متن در ایجاد این انگیزه مؤثرتر است. علاوه بر آن جنسیت و گروه سنی کاربران از پارامترهای مؤثر در اثرپذیری از محتوا در شبکه‌های اجتماعی شناخته شد. در مطالعه افراخته و همکاران (۱۳۹۵) از تحلیل شبکه‌های اجتماعی به منظور آشکارسازی ساختار حکمرانی محلی منابع آب استفاده شد. نتایج این بررسی نشان داد تمرکززدایی و واگذاری نقش‌های محوری به سازمان‌های تخصصی که زمینه‌ساز مدیریت مشارکتی منابع آب می‌باشد اتفاق نیفتاده و در مقابل تمرکز قدرت و اقتدار در مدیریت منابع آب کشاورزی در یک سازمان غیرتخصصی که نقش تسهیل‌کنندگی و بسترسازی دارد صورت گرفته است. پژوهشگران به منظور دستیابی به یافته‌های کارآمدتر پیشنهاد دادند علاوه بر روش تحلیل شبکه از روش‌های کیفی نظیر تحلیل محتوا، تئوری مبنا، مطالعات موردی نیز استفاده گردد.

۱۲- نقش کلان داده و شبکه‌های اجتماعی در کتابخانه‌ها

کتابخانه‌ها همواره بزرگ‌ترین فراهم‌کننده منابع علمی و تسهیل‌کننده دسترسی به این منابع بوده و هستند. پویایی کتابخانه‌ها با حرکت از الواح گلی به سمت کارت‌های فهرست‌نویسی و سپس به سمت دیجیتالی شدن صورت گرفته و مهم‌ترین نقش را در تهیه اطلاعات به‌منزله یک رسانه اجتماعی داشته‌اند. علاوه بر نقش سنتی کتابخانه‌ها، امروزه کتابداران از پذیرندگان فناوری‌های جدید به‌خصوص شبکه‌های اجتماعی به منظور ارتباط با کاربران هستند. گرچه کتابخانه‌ها همیشه از طریق اطلاعات با افراد در ارتباط هستند، رسانه‌های اجتماعی امروزه نقش مهمی برای کاربران کتابخانه در بازاریابی اطلاعات ایفا می‌کنند و کتابخانه را قطب بزرگ اجتماعی معرفی می‌کنند (فرناندز[1] ۲۰۰۹). با گسترش استفاده همگانی از شبکه جهانی وب و رشد فزاینده امکانات تعاملی آن برای به اشتراک‌گذاری اطلاعات و نیز تازه‌ها و رخدادها، گرایش متخصصان کتابداری و اطلاع‌رسانی به این قلمرو نیز افزایش یافته است. فعالیت متخصصان در

1- Fernandez

[Downloaded from jnrihs.ir on 2026-11-18]

فضای مجازی و رسانه‌های اجتماعی به سبب ویژگی‌ها، جذابیت‌ها و تأثیرگذاری این فضا در جامعه علمی و حرفه‌ای در حال گسترش است (زارعی و بیات، ۱۳۹۴).

با توجه به رشد تولیدات اطلاعاتی و نیاز کتابخانه‌ها به همگام شدن بیشتر با نیازهای کاربران و پاسخگویی به آن‌ها ضروری است تا کتابخانه‌ها با آخرین تحولات و پیشرفت‌های فناوری همگام شده و همواره خدمتشان را روزآمد سازند. برای این منظور یکی از ابزارهایی که بتواند این نیازها را پوشش دهد، شبکه‌های اجتماعی است. ضرورت به‌کارگیری چنین ابزارهایی به دلیل برخورداری از ویژگی‌هایی مانند چندرسانه‌ای بودن هم‌زمان، امکان برقراری تعامل بیشتر بین کاربران و کتابخانه، برقراری ارتباط سریع‌تر کاربران با کتابخانه و بالعکس و دسترسی سریع کاربران به اطلاعات موردنیاز بدون محدودیت مکان و زمان در کتابخانه‌ها برای ارائه خدمات بهتر و مطلوب‌تر بیش از پیش احساس می‌شود.

توییتر و فیس‌بوک معمول‌ترین برنامه‌های رسانه‌های اجتماعی هستند که به‌وسیله کتابخانه‌ها استفاده می‌شوند (پالمر[1] ۲۰۱۴) که بیانگر استفاده از توییتر و فیس‌بوک به عنوان محبوب‌ترین رسانه‌های اجتماعی که مورد استفاده عموم مردم هستند، تلقی می‌شود. یکی از ویژگی‌های کلیدی این ابزارها این است که امکان ارتباط دوطرفه و تعامل بین کتابخانه و کاربرانش را مقدور می‌سازند. این نوع ارتباط، معمولاً کارمندان و کاربران کتابخانه را دربرمی‌گیرد که درباره کتاب‌های موجود در مجموعه بحث می‌کنند و یا کارمندان را که پرسش‌های عمومی مرجع پاسخ می‌دهند. (بوتنگ و همکاران[2] ۲۰۱۴)؛ بنابراین می‌توان گفت رسانه‌های اجتماعی ابزارهای ارتباطی هستند که تنها از طریق آن‌ها اطلاعات کسب نمی‌شود، بلکه در آن‌ها تعاملی بین دریافت‌کننده اطلاعات و محمل اطلاعات به وجود می‌آید و این تعامل از یک سؤال ساده تا یک مقاله یا موضوع پیچیده می‌تواند ادامه یابد (زارعی و بیات، ۱۳۹۴).

با استفاده از روش‌های تحلیل مبتنی بر کلان داده‌های شبکه‌های اجتماعی می‌توان به بررسی محتواهایی پرداخت که توسط کاربران و یا کتابداران در شبکه‌های اجتماعی ارسال می‌شود و از طریق آن‌ها می‌توان به برآورده کردن نیازهای کاربران و جلب نظر مؤثر آن‌ها در استفاده از خدمات کتابخانه پرداخت. ضمن این‌که با بررسی نتایج تحلیل، به مدیران کتابخانه‌ها در روزآمد کردن سیستم و تصمیم‌گیری در ارائه خدمات بهتر، کمک شایان توجهی خواهد شد.

۱۳- نتیجه‌گیری

شبکه‌های اجتماعی با داشتن صدها میلیون کاربر در سراسر جهان، فرصتی باور نکردنی را برای ایجاد ارتباطات اجتماعی، یادگیری، تغییرات سیاسی و اجتماعی و سرگرمی‌های فردی و بالا بردن مهارت‌های گوناگون در اختیار افراد قرار می‌دهند. ازآنجاکه بسیاری از تعاملات اجتماعی امروزه در شبکه‌های

1- Palmer
2- Boateng et al

[Downloaded from jnrihs.ir on 2026-11-18]

اجتماعی اتفاق می‌افتد، دانشمندان علوم اجتماعی می‌توانند از این طریق به مقادیر بی‌سابقه‌ای از اطلاعات دسترسی داشته باشند. این نوع داده‌ها به علت حجم فراوان، تنوع گوناگون به صورت متن، تصویر، صوت، ویدئو و. . . و سرعت ایجاد فوق‌العاده در طبقه‌بندی کلان داده‌ها قرار می‌گیرند. برای تجزیه و تحلیل این‌گونه داده‌ها نیاز به روش‌های جایگزین روش‌های سنتی است تا به درستی استخراج، تحلیل و پردازش شوند و بتوانند به بهبود خدمات و کاهش هزینه‌ها کمک کنند. در کتابخانه‌ها نیز با توجه به رشد تولیدات علمی و تغییر نیازهای این مراکز، استفاده از شبکه‌های اجتماعی به منظور تعامل بیشتر با کاربران، شناسایی نیازها و تلاش در جهت رفع آن‌ها، ضروری به نظر می‌رسد و استفاده از روش‌های تحلیل شبکه‌های اجتماعی مبتنی بر کلان داده‌های موجود در آن‌ها در این حوزه می‌تواند در ارتقا، بهبود و توسعه خدمات و منابع کتابخانه مؤثر باشد.

[Downloaded from jnrihs.ir on 2026-11-18]

فهرست منابع و مآخذ

ـ یوسفی کندول، وحید. (۱۳۹۷). چهارچوب کاری برای پردازش‌های عظیم، کنفرانس بین‌المللی برق، کامپیوتر و مکانیک ایران. تهران: دبیرخانه دائمی کنفرانس
https://www.civilica.com/Paper-ECMCONF01-ECMCONF01_060. html

ـ حسینی، منیره و قلمکاری، افسون. (۱۳۹۶). تحلیل جوامع برند آنلاین بر پایه‌ی شبکه‌های اجتماعی با روش شبکه نگاری. سومین کنفرانس بین‌المللی وب‌پژوهی. تهران: دانشگاه علم و فرهنگ
https://www.civilica.com/Paper-IRANWEB03-IRANWEB03_026. html

ـ درمنده، مژده؛ نوروزی، علیرضا و سلطانی، افسانه. (۱۳۹۵). داده‌های بزرگ در شبکه‌های اجتماعی برخط: از تحلیل تعاملات کاربران تا مدل رفتار کاربران در شبکه‌های اجتماعی. اولین کنفرانس بین‌المللی بازیابی تعاملی اطلاعات. کیش: دانشگاه تهران
https://www.civilica.com/Paper-IIIRC01-IIIRC01_020. html

ـ ذوالفقارنسب، عاطفه. (۱۳۹۶). متن‌کاوی مقالات رشته کتابداری و اطلاع‌رسانی نمایه شده در پایگاه وب آو ساینس به منظور کشف ساختار درونی رشته و اعتبارسنجی آن. پایان‌نامه کارشناسی ارشد. دانشگاه علوم پزشکی و خدمات بهداشتی، درمانی کرمان.

ـ رضوانی، هومن و کبیری، شهاب. (۱۳۹۶). استفاده از تکنیک‌های داده‌کاوی در تحلیل اثرگذاری محتوای ترافیکی در شبکه‌های اجتماعی بر فرهنگ ترافیک. هفدهمین کنفرانس بین‌المللی مهندسی حمل‌ونقل و ترافیک. تهران: معاونت و سازمان حمل‌ونقل ترافیک
https://www.civilica.com/Paper-TTC17-TTC17_120. html

ـ زارعی، عاطفه و بیات، محمدکریم. (۱۳۹۴). کارایی رسانه‌های اجتماعی در کتابخانه‌های دانشگاهی: مطالعه موردی دانشگاه‌های دولتی تهران. کتاب مهر. ۱۳۹۴(۷-۸): ۹۸-۱۲۱.

ـ افراخته، حسن؛ طهماسبی، اصغر و عسکریه بزایه، فاطمه. (۱۳۹۵). کاربرد تحلیل شبکه اجتماعی در آشکارسازی ساختار حکمرانی محلی منابع آب. سومین همایش ملی پژوهش‌های مدیریت و علوم انسانی در ایران. تهران: موسسه پژوهشی مدیریت مدبر
https://www.civilica.com/Paper-IICMO03-IICMO03_320. html

ـ افسری شولی، فاطمه؛ هارون‌آبادی، علی و شامحمدی، امین. (۱۳۹۵). تحلیلی از داده‌های عظیم در شبکه‌های اجتماعی. دومین کنفرانس ملی رویکردهای نوین در مهندسی کامپیوتر و برق. رودسر: باشگاه پژوهشگران جوان و نخبگان
https://www.civilica.com/Paper-BPJ02-BPJ02_222. html

ـ زاهد، عطیه و سخی، محمدرضا. (۱۳۹۵). مروری بر تکنیک‌های کلان داده (Big Data). سومین کنفرانس ملی مهندسی برق و کامپیوتر سیستم‌های توزیع شده و شبکه‌های هوشمند. کاشان: دانشگاه آزاد اسلامی واحد کاشان
https://www.civilica.com/Paper-ELECONFK03-ELECONFK03_129. html

[Downloaded from jnrihs.ir on 2026-11-18]

ـ سنمی علمداری، یعقوب. (۱۳۹۵). مروری بر کلان داده‌ها BIG DATA. اولین همایش ملی نگرشی نوین در مهندسی برق و کامپیوتر. کرمانشاه: دانشگاه آزاد اسلامی واحد کرمانشاه

https://www.civilica.com/Paper-OUTLOOKECE01-OUTLOOKECE01_152. html

ـ دستاوردهای نوین در علوم مهندسی و پایه. تهران: مرکز پژوهش‌های زمین کاو

https://www.civilica.com/Paper-AEBSCONF01-AEBSCONF01_114. html

ـ علیزاده، فریبا؛ زارع، بابک و آقازاده، فهیمه. (۱۳۹۵). مروری بر مفهوم کلان داده. کنفرانس بین‌المللی مهندسی کامپیوتر و فناوری اطلاعات. تهران: دبیرخانه دائمی کنفرانس

https://www.civilica.com/Paper-CITCOMP01-CITCOMP01_312. html

ـ میرعابدینی، سیدجواد و روشن ضمیر، اکبر. (۱۳۹۵). تأثیر رایانش ابری بر توسعه کسب‌وکارهای کوچک و متوسط. دومین کنفرانس بین‌المللی مدیریت و علوم انسانی. استانبول- ترکیه: موسسه مدیران ایده پرداز پایتخت ویرا

https://www.civilica.com/Paper-ICMHCONF02-ICMHCONF02_178. html

ـ سهرابی، بابک و ایرج، حمیده. (۱۳۹۴). مدیریت کلان داده‌ها در بخش‌های خصوصی و عمومی. تهران: سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت).

ـ شریفی، مریم و سمیعی، اعظم. (۱۳۹۲). بررسی Big Data در شبکه‌های اجتماعی. اولین همایش ملی رویکردهای نوین در مهندسی کامپیوتر و بازیابی اطلاعات. رودسر: دانشگاه آزاد اسلامی واحد رودسر و املش

https://www.civilica.com/Paper-BPJ01-BPJ01_398. html

ـ مقدس، محمدصادق و آسا، برزین. (۱۳۹۲). آشنایی با رایانش ابری، ویژگی‌ها، نیازمندی‌ها، چالش‌ها. اولین کنفرانس ملی نوآوری در مهندسی کامپیوتر و فناوری اطلاعات. تنکابن: موسسه آموزش عالی شفق

https://www.civilica.com/Paper-CEIT01-CEIT01_119. html

ـ مولایی، محمدمهدی. (۱۳۹۲). سناریوهای آینده کلان داده. دنیای کامپیوتر و ارتباطات. ۱۴(۱۳۱): ۴۹-۳۸.

ـ مؤمنی راد، اکبر. (۱۳۹۲). تحلیل محتوای کیفی در آیین پژوهش: ماهیت، مراحل و اعتبار نتایج. اندازه‌گیری تربیتی. ۴(۱۴): ۱۸۷-۲۲۲.

ـ اسلامی، مروارید. (۱۳۹۱). بررسی شبکه‌های اجتماعی و تأثیرات آن‌ها بر ابعاد مختلف زندگی. نخستین کنگره ملی فضای مجازی و آسیب‌های اجتماعی نوپدی. تهران: وزارت تعاون، کار و رفاه اجتماعی

https://www.civilica.com/Paper-NOPADID01-NOPADID01_018. html

ـ بهشتی، محمدتقی و سروری، معین. (۱۳۹۱). رایانش ابر: ساختار، مزایا و چالش‌ها. اولین کارگاه ملی رایانش ابری. تهران: دانشگاه صنعتی امیرکبیر

https://www.civilica.com/Paper-CLOUDCOMPUTING01-CLOUDCOMPUTING01_006. html

[Downloaded from jnrihs.ir on 2026-11-18]

__ Singh, A. , Shukla, N. , & Mishra, N. (2018). Social media data analytics to improve supply chain management in food industries. Transportation Research Part E: Logistics and Transportation Review, 114, 398-415.
__ Sun, Z. (2018). 10 Bigs: Big Data and Its Ten Big Characteristics. Retrieved from: https://www. researchgate. net/profile/Zhaohao_Sun/publication/322592851_10_Bigs_Big_Data_and_Its_Ten_Big_Characteristics/links/5a6192cc4585158bca4a0a14/10-Bigs-Big-Data-and-Its-Ten-Big-Characteristics. pdf
__ Moessner, M. , Feldhege, J. , Wolf, M. , & Bauer, S. (2018). Analyzing big data in social media: Text and network analyses of an eating disorder forum. International Journal of Eating Disorders. 51(7), 656-667.
__ Mohammadpoor, M. , & Torabi, F. (2018). Big Data analytics in oil and gas industry: An emerging trend. Petroleum. Doi: https://doi. org/10. 1016/j. petlm. 2018. 11. 001
__ Stieglitz, S. , Mirbabaie, M. , Ross, B. , & Neuberger, C. (2018). Social media analytics–Challenges in topic discovery, data collection, and data preparation. International Journal of Information Management, 39, 156-168.
__ Nakatsugawa, M. (2018). Using big data analytics to advance precision radiation oncology. International Journal of Radiation Oncology* Biology* Physics, 101(2), 285-291.
__ Birjali, M. , Beni-Hssane, A. , & Erritali, M. (2017). Analyzing Social Media through Big Data using InfoSphere BigInsights and Apache Flume. Procedia computer science, 113, 280-285.
__ Kim, S. J. , Marsch, L. A. , Hancock, J. T. , & Das, A. K. (2017). Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data. Journal of medical Internet research, 19(10).
__ Kumari, S. (2016). Impact of big data and social media on society. Global Journal for Research Analysis. 5(3): 437-438.
__ Global social media research summary 2016. Retrieved from http://www. smartinsights. com/social-media marketing/socialmedia-strategy/new-global-social-media-research/
__ Hashem, I. A. T. , Yaqoob, I. , Anuar, N. B. , Mokhtar, S. , Gani, A. , & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. Information systems, 47, 98-115.
__ Boateng, F. , & Quan Liu, Y. (2014). Web 2. 0 applications' usage and trends in top US academic libraries. Library Hi Tech, 32(1), 120-138.
__ Chen, C. P. , & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275, 314-347.
__ Chen, M. , Mao, S. , & Liu, Y. (2014). Big data: A survey. Mobile networks and applications, 19(2), 171-209.
__ Palmer, S. (2014). Characterizing university library use of social media: a case study of Twitter and Facebook from Australia. The Journal of Academic Librarianship, 40(6), 611-619.

[Downloaded from jnrihs.ir on 2026-11-18]

__ Dhawan, V. , & Zanini, N. (2014). Big data and social media analytics. Research Matters. Retrieved from: http://www. cambridgeassessment. org. uk/Images/465808-big-data-and-social-media-analytics. pdf

__ Campbell, W. M. , Dagli, C. K. , & Weinstein, C. J. (2013). Social network analysis with content and graphs. Lincoln Laboratory Journal, 20(1), 61-81.

__ Center, I. I. (2012). Planning guide: Getting started with hadoop. Steps IT Managers can take to move forward with big data analytics.

__ Gantz, J. , & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC iView: IDC Analyze the future, 2007(2012), 1-16.

__ Gerhardt, B. , Griffin, K. , & Klemann, R. (2012). Unlocking value in the fragmented world of big data analytics. Cisco Internet Business Solutions Group.

__ Madden, S. (2012). From databases to big data. IEEE Internet Computing, (3), 4-6.

__ Minelli, M. , Chambers, M. , & Dhiraj, A. (2012). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. New Jersey: John Wiley & Sons.

__ Rao, B. P. , Saluia, P. , Sharma, N. , Mittal, A. , & Sharma, S. V. (2012). Cloud computing for Internet of Things & sensing based applications. 2012 Sixth International Conference on Sensing Technology (ICST). IEEE.

__ Singh, S. and Singh, N. (2011) Big Data Analytics. International Conference on Communication, Information & Computing Technology. Mumbai India: IEEE, October 2011.

__ Gantz, J. , & Reinsel, D. (2011). Extracting value from chaos. IDC iview, 1142(2011), 1-12.

__ Fernandez, J. (2009). A SWOT analysis for social media in libraries. Online, 33(5), 35.

__ McNutt, T. R. , Benedict, S. H. , Low, D. A. , Moore, K. , Shpitser, I. , Jiang, W& Mell, P. , &

__ Grance, T. (2009). Effectively and securely using the cloud computing paradigm. NIST, Information Technology Laboratory, 2(8), 304-311.

[Downloaded from jnrihs.ir on 2026-11-18]