تشخیص k پست اثرگذار برمبنای یادگیری عمیق در جهت بهبود مدیریت دانش

حسنیه صفی آریان[۱]، محمدجعفر تارخ[۲،*]، محمدعلی افشار کاظمی[۳]

[۱]دانشجوی دکتری، گروه مدیریت، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
[۲]استاد، گروه صنایع، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران (عهده‌دار مکاتبات)
[۳]دانشیار، گروه مدیریت، واحد تهران مرکز، دانشگاه آزاد اسلامی، تهران، ایران
تاریخ دریافت: اردیبهشت ۱۴۰۱، اصلاحیه: مرداد ۱۴۰۱، پذیرش: مهر ۱۴۰۱

چکیده

حضور رسانه‌های اجتماعی باعث ایجاد تحولات اساسی در جوامع امروزی شده است. این رسانه‌ها بستری مناسب برای کاربران در جهت اشتراک‌گذاری دانش محسوب می‌شوند. همین امر موجب شده است که کاربران با انبوهی از اطلاعات مواجه شوند که در بیشتر مواقع مناسب آن‌ها نباشد و نفوذ کم‌تری بر روی آن‌ها داشته باشد، در نتیجه اطلاعات و دانشی که بین کاربران به اشتراک گذاشته می‌شود کارآمد نخواهد بود و چرخه مدیریت دانش به درستی بر روی این رسانه‌های اجتماعی اجرا نشود. ارائه روشی جهت انتخاب پست‌های اثرگذار برای کاربر در میان انبوهی از پست‌ها، می‌تواند بسیار حائز اهمیت باشد. روش‌هایی که در پژوهش‌های اخیر در جهت انتخاب پست‌های اثرگذار ارائه شده است مبتنی بر خصیصه‌های آماری مربوط به داده‌های مختلف میکروبلاگ‌ها می‌باشند و کمتر به صورت محتوایی اثرگذاری، هر پست را بر روی کاربر مشخص، مورد اندازه‌گیری قرار داده است. علی‌رغم تنوع موضوعی، محتوایی توئیت‌ها و کاربران مختلف، اکثر این روش‌ها با ارائه یک مدل عمومی بر مبنای خصیصه‌های پر تعداد، از دقت برخوردار نیستند و قادر به ارائه پیشگویی در زمان برخط نمی‌باشند. در این پژوهش، با تحلیل انتشار پست‌ها بین کاربران در بازه زمانی مشخص، به بررسی روشی برای سنجش توجه کاربران به مطالب به اشتراک گذاشته شده و تأثیرات آن‌ها پرداخته می‌شود، این روش **IKS** نام گذاری شده است که بر مبنای خصیصه‌های محتوای منتشر شده توسط کاربر ارائه شده است و به صورت یک مساله کلاس‌بند دودوئی که بر مبنای یادگیری عمیق می‌باشد، طراحی شده است. ارزیابی این روش با استفاده از روش شهودی و ارزیابی مجموعه دادگان انجام شده است که دقت بیشتری در مقایسه با سایر روش‌ها دارد.

واژه‌های اصلی: انتشار، پست اثرگذار،کاربر ،مدیریت دانش، یادگیری عمیق

۱- مقدمه

مروزه، محققان بر روی نقش دانش در دستاوردهای مزیت رقابتی برای ظهور و پیشرفت دیدگاه دانش در بازاریابی متمرکزشده‌اند[۳۶]. دانش، جایگزین دستگاه‌های مکانیکی، سرمایه و مواد خام شده است تا منابع تولیدی مهمی را برای سازمان‌ها ایجاد کند. دانش علاوه بر ایجاد شود برای سازمان‌ها با غنی شدن آن به صورت مستمر و تطبیق شدن محتوای آن در پایگاه‌های دانشی، ارزش دانش را افزایش می‌دهد. از طرفی افزایش رقابت بین شرکت‌ها باعث شده است مدیریت دانش و حساسیت بازار را مهم جلوه دهد[۱۵]. بحث روی مدیریت دانش باعث بهبود شناخت بر روی اهمیت دانش در (اقتصاد دانشی جدید) و تأثیر بر روی مزیت رقابتی سازمان‌ها می‌شود.

مدیریت دانش به عنوان فرآیندی است که، سازمان به تولید، بدست‌آوری، تسخیر و به کارگیری دانش برای ارتقاء بهره وری سازمان می‌پرازد، شناخته شده است[۱۵]. مدیریت دانش، شامل نگهداری، به اشتراک گذاشتن، بهبود دادن و مؤثر کردن دانش در محیط سازمانی می‌باشد[۳۷]. اهداف مدیریت دانش در سه سطح تاکتیکی، سیاست‌های توسعه‌ایی، اهداف کوتاه مدت مربوط به نیازهای اجرایی و در سطح عملیاتی، روندها، قوانین داخلی و کارهای روزمره مدنظر است[۳۶]. مدیریت دانش مستلزم این است که سازمان، محصولاتش را سازماندهی کرده و با مشتریان اعم از خریدارن و تأمین‌کنندگان مواد اولیه نه بر پایه مشخصات عمومی از پیش فرض شده بلکه بر مبنای ترجیحات واقعی مشتریان تعامل داشته باشد. در حالی که سازمان‌ها در راستای اهداف مدیریت ارتباط با مشتریان رو به جلو حرکت می‌کند، عملکرد بازاریابی به عنوان خطوط مقدم در تعامل با مصرف کننده و تأمین کننده نقش اساسی دارد[۲۶]. کاربرد مدیریت دانش در بازاریابی، کسب وکارها را توانمند کرده است که نوع جدیدی از نیازها را ایجاد کند که با نیازمندی‌های مصرف کننده امروز، انطباق بیشتری داشته باشد[۷]. مدیریت دانش یک استراتژی تجاری می‌باشد که به هسته سازمان‌ها در برابر تغییرات محیط بازار، بهبود رقابت و ایجاد هماهنگی ساختار داخلی سازمان کمک می‌کند.

* mjtarokh@kntu.ac.ir

شرکت‌ها تمایل دارند متقاعد کنند که آنها یک فهم قابل توجهی در ارتباط با مشتری‌هایشان و موقعیت بازار دارند، این شناخت نادرست باعث می‌شود مدیران تصمیمات اشتباهی بگیرند. در دیدگاه دوکر، فقط مشتری به درستی بازار را درک می‌کند بنابراین شرکت‌ها باید تحقیقات جامعی را در بازار انجام دهند و رفتار مشتری را در بازار تحلیل کنند و با پیشگویی نیازهای آینده مشتری از بازار غیرفعال به فعال تبدیل شوند[۴۱] .

در سال‌های اخیر حضور رسانه‌های اجتماعی باعث توسعه و تحول بسیار جدی در میان تعاملات انسانی شده است. کاربران این رسانه‌های اجتماعی به راحتی صفحات مورد علایق خودشان را پیگیری می‌کنند و اطلاعات مورد نظرشان را با کاربران دیگر به اشتراک می‌گذارند. یکی از مباحث بسیار مهم که در این حوزه مطرح می‌شود مدیریت دانش در این رسانه‌ها می‌باشد. به عبارتی دیگر، مدیریت دانش در بسیاری از اجتماعات برخط به عنوان یکی از روش‌های مطرح در زمینه تولید و اشتراک‌گذاری اطلاعات در نظر گرفته می‌شود. مدیریت دانش نقش مهمی در ثبت و استانداردسازی دانش‌های نهفته در گروه‌های مختلف کاربران دارد[۲۸] .

اجتماعات برخط تخصص محور به گروه‌های در زمینه‌های خاص گفته می‌شوند که در زمینه‌های معینی علاقه‌مندی مشترک دارند و اقدام به تولید و اشتراک‌گذاری دانش با یکدیگر می‌نمایند. اغلب این اجتماعات در بسترهای وب و یا تالارهای پرسش و پاسخ اقدام به انتقال یافته‌ها و دانش خود به صورت متن، فرامتن و چند رسانه‌ایی می‌نمایند. با گسترش نمونه‌های این گروه‌ها و تالارها نظیر yahoo answers!، Twitter و یا stack overflow و غیره، بحث مهم افزایش مشارکت در این گروه‌ها است. مشارکت بیشتر موجب گسترش اطلاعات، تبادل ایده‌ها و نوآوری‌ها در این گروه‌ها خواهد شد[۲۸]. اما موضوع اساسی در این بسترها این مورد است که حجم انبوه این اطلاعات باعث سردرگمی کاربران می شود و در بیشتر اوقات نمی توانند موضوعات مورد علایق خود را ببینند و با سایر کاربرانی که علایق مشترک دارند در میان بگذارند.

باتوجه به این ضرورت مساله، در این مقاله، با تحلیل انتشار پست‌ها بین کاربران در بازه زمانی مشخص، به بررسی روشی برای سنجش توجه کاربران به مطالب به اشتراک گذاشته شده و تأثیرات آن‌ها پرداخته می‌شود. در روش پیشنهادی این پژوهش، علاوه بر توجه به ویژگی‌های ساختاری کاربران در شبکه، به رفتار انتشار اطلاعات بین کاربران توجه می‌گردد. تأثیر یک کاربر، فقط به میزان کاربران پیگیری‌کننده آن بستگی ندارد، بلکه به میزان تاثیرپذیری کاربران نیز بستگی دارد. عمده روش‌های ارائه شده در این موضوع بر پایه شناسایی افراد موثر در انتشار اطلاعات در گروه می‌باشد. به این صورت پست‌ها اکثرا به افراد موثر نمایش داده می‌شود که در صورت بازنشر، تعداد بیشتری از اعضای گروه این پست‌ها را مشاهده نمایند. این روش‌ها عمدتا بر پایه ویژگی‌های آماری فردی نظیر تعداد کاربران پیگیری‌کننده و میزان بازنشر می‌باشند. با توجه به تاثیرپذیری کاربران در انتشار دانش در گروه مشتاقان، در این پژوهش روش پیشنهادی برای تعیین ارزش محتوایی توئیت‌ها، شامل معرفی و بکارگیری خصیصه‌هایی است که عمدتا به بررسی محتوایی یک توئیت و میزان اثرگذاری آن بر روی کاربر مورد نظر می‌پردازد، تا بتواند رفتار بازنشر این کاربر را در قبال این توئیت پیشگویی نماید. از جمله خصیصه‌های پیشنهادی برای تعیین ارزش محتوایی یک توئیت، تازگی(غیرمنتظره بودن) توئیت می‌باشد. میزان تازگی یک توئیت یک خصیصه نسبی برای هر کاربر بوده و متناسب با ارتباط موضوعی توئیت با توئیت‌های منتشرشده توسط کاربر می‌باشد. در این مقاله ضمن بیان روشی برای اندازه‌گیری میزان تازگی یک توئیت برای یک کاربر، به بررسی ارتباط این خصیصه با رفتار بازنشر کاربران پرداخته می‌شود. توئیت‌هایی با محتوای منفی، با توجه به[۹] از قابلیت بازنشر بیشتری برخوردار می‌باشند. منفی بودن یک توئیت خصیصه دیگری است که برای بازنشر معرفی و روشی برای اندازه‌گیری آن به صورت برخط ارائه می‌شود. شباهت موضوع یک توئیت با علایق کاربر (محتویات توئیت‌های قبلی کاربر) و ارتباط مکانی احتمالی توئیت با کاربر در تصمیم برای بازنشر کاربران بسیار موثر است. در کنار این خصیصه‌ها، خصیصه‌هایی مانند نرخ رشد موضوع یک توئیت و وزن اجتماعی یک توئیت، در این مقاله معرفی و برای اندازه‌گیری آن‌ها روش‌هایی پیشنهاد می‌گردد. به منظور پیشگویی از روش کلاس‌بندی باینری با استفاده از یادگیری به کمک شبکه عصبی عمیق استفاده شده است. یادگیری عمیق تحول شگرفی در افزایش دقت کلاس بند داشته است. برای این کار، مسئله به صورت یک مسئله یادگیری کلاس‌بند با استفاده از شبکه عصبی عمیق تعریف شده است. ارزیابی روش پیشنهادی به دو صورت شهودی و با آزمایش مجموعه داده صورت می‌گیرد. در ارزیابی شهودی با کمک گروه کاربران داوطلب و ارائه توئیت‌های روزانه، میزان همبستگی بین خصیصه‌های پیشنهادی با رفتار بازنشر با کمک ضریب همبستگی Point-biserial اندازه‌گیری می‌شود. همین طور دقت روش‌های اندازه‌گیری خصیصه‌های پیشنهادی با کمک ضریب همبستگی رتبه‌ای کاندال ارزیابی می‌گردد.

ادامه مقاله به این صورت می‌باشد که در قسمت بعد به مرورکارهای مرتبط به حوزه بازنشر اطلاعات و پیشگویی بازنشر پرداخته می‌شود. در قسمت ۳، ضمن معرفی خصیصه‌های پیشنهادی و روش اندازه‌گیری برای تخمین ارزش محتوایی هر توئیت برای هر کاربر، روش کلاس‌بندی باینری با ناظر با روش شبکه عصبی عمیق ارائه می‌گردد. برای ارزیابی روش پیشنهادی، در قسمت ۴، آزمایش با کمک مجموعه داده (سه مجموعه داده از توئیتر) برای ارزیابی انتخاب شده است. در قسمت ۵، جمع‌بندی مقاله و نتیجه گیری ارائه می‌گردد.

۲- کارهای مرتبط

در این بخش به ارائه کارهای مرتبط در معیارهای شبکه اجتماعی و افزایش انتشار در شبکه اجتماعی پرداخت می‌شود.

۲-۱- معیارهای شبکه اجتماعی

چند معیار مرتبط با تحلیل شبکه اجتماعی (SNA) وجود دارد که تأثیر

اجتماعی کاربران را در اجتماعات برخط نشان می‌دهند. سه معیار مرسوم به ترتیب ۱- مرکزیت درجه که تعداد ارتباطات مستقیم یا غیرمستقیم یک گره به دیگر گره‌ها را نشان می‌دهد، می‌باشد[۱۷] در[۲۳] کاربرانی که خوب به یکدیگر متصل شده‌اند را «محور» نامیده می‌شوند. ۲-مرکزیت نزدیکی در اینجا برخلاف مرکزیت، درجه تنها روابط مستقیم را مدنظر قرار می‌دهد و بر فاصله‌ی یک کاربر از دیگران در شبکه تأکید می‌کند، در واقع این معیار بر فاصله‌ی هر کاربر از همه‌ی کاربران دیگر تمرکز می‌کند. ۳-مرکزیت میانی: تعداد دفعاتی که یک کاربر به عنوان پل در کوتاه‌ترین مسیر میان دو کاربر نقش ایفا می‌کند را، مرکزیت میانی می‌نامند. در فضای مشابه، در پژوهش[۱۳] اشاره می‌کند که چنین پل‌هایی که دو بخش غیرمتصل از شبکه را به هم متصل می‌کنند، تأثیرگذار هستند.

۲-۲- مدل های انتشار در شبکه‌های اجتماعی

انتشار اطلاعات در شبکه‌های اجتماعی مشابه با انتشار بیماری است، اما تفاوت‌های بسیاری با یکدیگر دارند. در انتشار اطلاعات فاکتورهای از جمله، زمان انتشار، محتوای انتشار، کاربران، ارتباطات کاربران با یکدیگر و ساختار شبکه بسیار موثر است[۲۲] محققان مدل‌هایی که در بخش قبل اشاره شد را بهبود داده‌اند و در ادامه خلاصه‌ایی از این مدل‌ها به شرح زیرارئه خواهد شد:

در[۳۸]گره E به مدل SIR اضافه شد، و این مدل SEIR نامیده شده است. آن‌ها یک ارتقا تکاملی پویا را ایجاد کردند تا پردازش اطلاعات را توصیف کنند، و تاثیر آن بر روی فرکانس ورود کاربران و تعداد دوستان آنها در انتشار اطلاعات مورد بررسی قرار گرفت. نتایج نشان می‌دهد فرکانس ورود کاربران بر روی سرعت انتشار و انعکاس اطلاعات تاثیر مستقیمی می‌گذارد. در این مدل رفتارهای متفاوت کاربران نادیده گرفته شده است و این یک نکته کلیدی بوده است که در انتشار اطلاعات می‌تواند نقش مهمی را ایفا کند، بنابراین این مدل از دقت کافی برخوردار نیست.در[۴۱] انتشار اطلاعات به رفتار کاربر تنها وابسته نیست اما این اطلاعات می‌تواند حائز اهمیت باشند بنابراین این تیم یک مدلی به نام S-SEIR ارائه داد که بر پایه SEIR است. در[۸] یک مدلی به نام SCIR برای میکروبلاگ‌ها با گزینه ارتباطات ایجاد کردند در این مدل فرض بر این است که تمام فن‌ها به گزینه ارتباط متصل خواهد شد زمانی که یک پیام منتشر خواهد شد. بنابراین روند فن‌ها بنا به احتمالات خاصی در آینده تغییر خواهد کرد و می‌تواند باعث انتشار اطلاعات در یک مدت زمان مشخص باشد. این مدل می‌تواند به طور منطقی انتشار موضوعات را نشان دهد. در [۱۹] یک مدل به نام isSIR برپایه مدل SIR ایجاد کردند که فرآیند انتشار عفونت را در این مدل شبیه‌سازی کرده‌اند.که این مدل می‌تواند در گوگل و فیس بوک بسیار متنوع باشد. انتشار اطلاعات متفاوت از اپیدمی بیماری است[۱۰]. حد آستانه‌ایی برای گرفتن بیماری واگیردار وجود ندارد در حالی‌که این حد آستانه برای دریافت اطلاعات موجود است. در این مدل اشاره می‌شود به این موضوع که اگر فردی با حجم بزرگی از اطلاعات مواجه شود ممکن است تمایلی به خواندن آن‌ها نداشته باشد و در نتیجه تلاشی به انتشار آن‌ها در شبکه‌های اجتماعی نکند. بنابراین آن‌ها مدل FSIR را ارائه دادند تا بر همسایگان و افراد حاضر در شبکه برای انتشار اطلاعات تاکید کنند. مدل EIS را براساس مدل SIS پیشنهاد دادند[۳۰] در این مدل وقتی اطلاعاتی بین کاربران منتشر می‌شود در اصل یک پیام احساسی بین کاربران رد و بدل می‌شود. در این مدل انتشاراطلاعات مرتبط با احتمال انتشار و چگالی انتقال است و بنابراین عملکردی بهتر از SIS دارد.

۲-۳ - افزایش انتشار در شبکه‌های اجتماعی

در پژوهش[۶]، با استفاده از دو درگاه اشتراک‌گذاری محتوای یوتیوب و Digg، نشان داده شده است که، چگونه با نظارت بر پاسخ به روایت‌ها، می‌توان تصمیم باز نشر چنین روایت‌هایی را با دقت قابل توجهی پیش‌بینی کرد. در مطالعه‌ی دیگر در[۲۰]، نقش اجتماعات بر خط در محتوای بازنشر شده Digg، مورد بررسی قرار گرفته است. آن‌ها فهمیدند که الگوهای پخش علاقه در یک روایت در اجتماع برخط، چگونگی محبوب شدن این روایت را نشان می‌دهد. در حوزه‌ی دیگر، در[۲] آبشار گردش اطلاعات شبکه‌های بزرگ، به صورت یک مجموعه بزرگ از پیشنهادهای «کاربر-کاربر» در نظر گرفته شده است و الگوهای آبشاری بازنشر در اجتماعات بزرگ برخط، پیش‌بینی شده‌اند. معیارهای اصلی دیگری در[۲] اضافه شده‌اند که می‌توانند در پیش‌بینی تصمیم بازنشر، مؤثر واقع شوند. این عوامل عبارت‌اند از (الف) روابط میان فردی بین کاربران معمولی، (ب) آمادگی یک جامعه برای پذیرش یک پست. این نگاه جدید بر روی تصمیم بازنشر، منجر به ارائهٔ بسیاری از راهبردهای بازاریابی از جمله تصفیه همکارانه در اجتماعات برخط شد. این راهبرد روشی است که در برخی سامانه‌های توصیه‌گر اجتماعات برخط، استفاده می‌گردد.

درک چگونگی ایجاد یک پست توسط کاربران، و انگیزه‌های آن‌ها برای بازنشر یک پست، در پیش‌بینی تصمیم بازنشر یک پست، دارای اهمیت فراوان دارد. در واقع، فهمیدن این‌که کاربران چه محتواهایی را برای بازنشر انتخاب می‌کنند، می‌تواند به توضیح چرایی بازنشر یک پست خاص و پیش‌بینی آن، کمک کند. این انگیزه‌ها برای تصمیم بازنشر در مطالعه‌ی انجام شده در[۱] به‌خوبی بررسی شده‌اند. آن‌ها دلایل اصلی تصمیم بازنشر توسط کاربران، را مشخص نموده و ده انگیزه‌ی مختلف را برای بازنشر معرفی کرده‌اند. از جمله این انگیزه‌ها می‌توان به نظر دادن در پست‌ها، انتشار پست‌ها به گیرندگان جدید برای آگاهی دادن به اشخاص یا گروه‌های خاص و ذخیره‌ی پست‌ها برای دسترسی آتی شخصی را، اشاره کرد. اگرچه تمرکز مطالعه‌ی آن‌ها پیش‌بینی تصمیم بازنشر پست‌ها نبوده است، اما انگیزه‌های گفته‌شدهٔ برای تصمیم بازنشر، می‌تواند در تعیین ویژگی‌ها مؤثر برای پیش‌بینی تصمیم بازنشر پست‌ها، مورد استفاده قرار گیرند.

مطالعات دیگری برای یافتن عوامل بازنشر توسط سو و همکارانش در [۱۸] و [۲۷] انجام شده است. آن‌ها سه عامل نهفته از ویژگی‌های پست

را، با استفاده از روش تحلیل مؤلفه‌ی اصلی (PCA) استخراج کرده و تلاش کردند تا آن‌ها را به ویژگی‌های واقعی آشکار، ارتباط دهند. آن‌ها سپس یک مدل خطی برای یافتن احتمال بازنشر معرفی کردند. آن‌ها انگیزه‌ی خود را در انتخاب مدل خطی برای پیش‌بینی تصمیم بازنشر، ذکر نکرده‌اند و درباره‌ی مؤثر بودن روش PCA در یافتن عوامل مهم قابلیت بازنشر نیز بحث ننموده‌اند. همچنین آن‌ها تنها آزمایش‌های خود را بر روی مجموعهٔ محدودی از داده‌ها که خودشان ارائه دادند، اجرا کردند. آن‌ها نتیجه گرفتند که ویژگی‌های مبتنی بر محتوا نظیر هَشتَگ و url در قابلیت بازنشر، بسیار مهم هستند. البته این نتیجه‌گیری توسط مطالعات بعدی در[۲۹]، به چالش کشیده شد. آن‌ها نشان دادند که ویژگی‌های مربوط به محتوای یک پست به‌تنهایی، دارای اطلاعات کافی برای پیش‌بینی تصمیم بازنشر پست‌ها نمی‌باشند.

در یک مطالعه‌ی مشابه[۱۶]، یک تحقیق تجربی برای پیش‌بینی تصمیم بازنشر یک پست انجام شد. در این پژوهش، یک الگوریتم مبتنی بر یادگیری برخط، برای پیش‌بینی تا حد ممکن سریع، توسعه داده شد. برای مطالعات بیشتر در ارتباط با الگوریتم‌های مبتنی بر یادگیری برخط به مقاله[۳۴] مراجعه نمایید. سپس، برای مرحله آموزش مدل‌های ارائه شده، از زیرمجموعه‌های مختلف داده که بر اساس زمان و روز تولید شده بودند، استفاده کردند تا از اطلاعات زمانی پست‌ها برای پیش‌بینی بازنشر یک پست، بهتر بهره ببرند. در این پژوهش، همانند مطالعه‌ی[۱۳]، دلیل استفاده از این مدل به صراحت بیان نشده است و مدل با استفاده از مجموعه دادگان مختلف، مورد ارزیابی قرار نگرفته است. در این پژوهش، عملکرد روش یادگیری برخط، با پیش‌بینی‌های انسانی مقایسه شده و ادعا شده است که این روش به‌خوبی پیش‌بینی‌های انسانی عمل می‌کند. زَمان و همکارانش در[۲۴] نیز، پژوهشی در ارتباط با پیش‌بینی تصمیم بازنشر بر اساس روش تصفیه همکارانه انجام دادند. برخلاف مطالعات دیگر که از ویژگی‌های مستقیماً استخراج شده از پست‌ها یا کاربران استفاده می‌کنند، آن‌ها از بازخوردهای مثبت و منفی غیرمستقیم در مدلشان استفاده کردند. اگر کاربران پیگیری‌کنندهٔ فعال یک پست را بازنشر کنند، به‌عنوان بازخورد مثبت، و در غیر این صورت به عنوان بازخورد منفی در نظر گرفته می‌شود. اما یک نقطه ضعف در این پژوهش این است که آن‌ها مدل‌های خود را بر اساس حداقل یک ساعت داده‌ی پس از انتشار یک پست آموزش دادند. از سوی دیگر، مطالعات قبلی نشان می‌دهند که بیشتر از نود درصد پست‌های بازنشر شده، در اولین ساعت پس از ایجاد اتفاق می‌افتند. بنابراین، آموزش یک مدل بر اساس چنین بازه‌ی زمانی طولانی از نظر عملی ارزشمند نیست.

آرتزی و همکارانش در[۴۴]، یک مدل متمایزکننده را برای پیش‌بینی احتمالاتی بازنشر یک پست، توسعه دادند. آن‌ها چندین ویژگی تاریخچه‌ای و لغوی از متن پست، و برخی ویژگی‌ها را از کاربران انتشاردهندهٔ پست‌ها، استخراج کردند. آن‌ها برای پیش‌بینی از دو طبقه‌بندی مختلف به نام درخت رگرسیون تجمیعی چندگانه[۳] و طبقه‌بندی بیشین آنتروپی[۳۹] استفاده کردند. تمرکز اصلی مطالعه‌ی آن‌ها بر روی ویژگی‌های محتوایی بود. بنابراین، به عنوان یک محدودیت، کار آن‌ها تنها بر روی پست‌های انگلیسی انجام شد. همچنین آن‌ها از ویژگی‌هایی نظیر هَشتَگ که برای پیش‌بینی تصمیم بازنشر پست‌ها مفید است، استفاده نکردند. در[۲۵]، از روش تخصیص پنهان دریکله برای پیش‌بینی رفتار بازنشر کاربران استفاده شده است. این روش در محیط‌های برخط قابلیت عملیاتی ضعیفی دارد.

در[۲] مشابه[۴۵]، یک روش مبتنی بر شبکه‌های بِیزی، برای پیش‌بینی تعداد بازنشرهای یک پست معلوم بر اساس الگوی پخش اولیهٔ آن، ارائه شده است. در این پژوهش، مسئلهٔ پیش‌بینی تصمیم بازنشر بر اساس مطالعه‌ی الگوی پخش پست‌ها انجام شده است. آن‌ها دریافتند که زمان‌های واکنش به پست‌ها را می‌توان به خوبی با استفاده از یک توزیع نُرمال لگاریتمی تخمین زد. در این پژوهش، از شبکه بیزی برای مدل‌سازی پویایی بازنشرها در طی زمان، استفاده شده است. برخلاف سایر تحقیقات که عموماً در آن‌ها مدل‌های یادگیری ویژگی‌محور هستند، این تحقیق هیچ ویژگی را به شکل مستقیم از پست‌ها یا کاربر دریافت نمی‌کند. در واقع پیش‌بینی تنها بر پایه‌ی الگوهای انتشار قبلی پست‌ها بنا شده است. آن‌ها ادعا می‌کنند که راهکارشان زمانی خوب کار می‌کند که حداقل ۱۰٪ بازنشرهای یک پست، دیده شوند. این موضوع برای ما خیلی جالب نیست چرا که اولاً روشن نیست چه زمانی ۱۰٪ پست‌ها مشاهده می‌گردد، و ثانیاً در مسئله مطرح شده در این رساله، پیش‌بینی بازنشر یک پست پیش از انتشار و یا مدت کمی پس از انتشار آن‌ها مورد نظر است.

در یک تحقیق دیگر[۴]، یک مدل وزن‌دهی شده بر اساس ویژگی ارائه شد که تصمیم بازنشر پست‌ها را بر اساس تعداد بازنشرهای بالقوه، پیش‌بینی می‌کند. برخلاف کارهای دیگر، این کار یک عمل کلاس‌بندی چندگانه است که در آن یک پست، در یکی از چهار کلاس ممکن جای می‌گیرد. این کلاس‌ها عبارتند از کلاس صفر یا پست‌های بازنشر نشده، کلاس یک یا پست‌های با کمتر از ده بار بازنشر، کلاس دو یا پست‌های با کم‌تر از ۱۰۰ بار بازنشر و کلاس سه یعنی پست‌های با بیش از ۱۰۰ بار بازنشر. مدل استخراج ویژگی آن‌ها، مجموعه‌ای از ویژگی‌ها را از خود پست و از کاربر منتشرکننده‌ی پست استخراج می‌کند. برخلاف[۴۴] تمرکز آن‌ها بیشتر بر روی ویژگی‌های اجتماعی است. آن‌ها از یک جداکنندهٔ ماشین بردار پشتیبان با هسته‌ی استفاده کردند که مدل آن‌ها را قادر می‌سازد مرزهای پیچیده‌ای برای تفکیک کلاس‌ها به وجود آورد. در روش آن‌ها که بر اساس وزن‌دهی بنا شده است، به هر ویژگی یک وزن اختصاص می‌دهد که بر اساس بهره اطلاعات هر ویژگی، محاسبه می‌شود. این روش به هر ویژگی که بهره اطلاعاتی بیشتری داشته باشد، وزن بالاتری اختصاص می‌دهد. این نوع وزن‌دهی به ویژگی‌ها، سبب می‌شود آن ویژگی‌ها سهم بیشتری در عمل کلاس‌بندی داشته باشند. مقادیر وزن بر اساس یک ارزیابی آزمایشی بر روی مجموعه دادگان، به دست آمده است. هرچند گزارش شده است که روش آن‌ها از روش‌های بدون استفاده از وزن‌دهی بهتر عمل می‌کند، و لیکن نویسندگان مدارک کافی ارائه نکرده‌اند تا نشان دهند این روش برای مجموعه دادگان دیگر و با تنظیمات دیگر نیز، بهینه است.

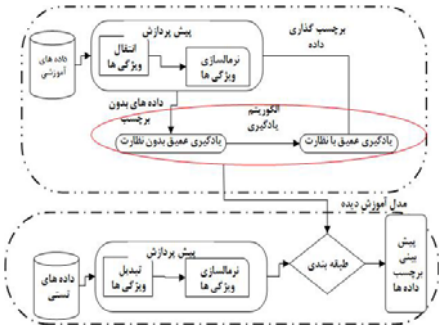
یک تحقیق مشابه برای پیش‌بینی میزان تصمیم بازنشر پست‌ها نیز توسط هونگ و همکاران[۴۲]، انجام گرفته است. آن‌ها این عمل را بر اساس دو مسئله‌ی کلاس‌بندی دودویی و چندگانه نظیم کردند. هدف در عمل کلاس‌بندی دودویی این است که مشخص کند یک پیام بازنشر خواهد شد یا خیر، و عمل کلاس‌بندی چندگانه سعی می‌کند حجم بازنشرهای یک پست را پس از انتشار آن پیش‌بینی کند. آن‌ها از روش TF-IDF برای پردازش ویژگی‌های محتوا استفاده کردند. ولی مشخص نکردند دقیقاً از چه رده‌بندی استفاده کرده‌اند و کدام ویژگی‌ها بیشتر در عمل کلاس‌بندی مشارکت داشته است. همچنین روش آن‌ها از نظر عمومیت نیز مشکل دارد، زیرا که بر روی مجموعه دادگان‌ی محدودی آزمایش شده است.

۲-۴- نقش شبکه‌های اجتماعی در مدیریت دانش

شبکه‌های اجتماعی باعث ایجاد ایده‌های جدید، قدرت تصمیم‌گیری در کاربران و تعامل برای تصمیم‌گیری بین آن‌ها ایجاد کرده است. همچنین فرهنگ جدیدی در سازمان‌ها برای اشتراک‌گذاری آزاد دانش و تشویق کاربران جهت این کار به وجود آورده است[۲۱]. در[۱۲] تاثیری که رسانه‌های اجتماعی بر روی مدیریت دانش در سازمان‌ها داشته است در ۴ رکن بررسی می‌شود، رکن اول آن توسعه مرزهای مدیریت دانش در سازمان‌ها است. رکن دوم تضعیف مرزهای سازمانی است. رکن سوم، تغییر ارتباط بین کارمندان است. رکن چهارم، توسعه تکنولوژی موبایلی در سازمان‌ها است. در[۳۲] بیان می‌کند رسانه‌های اجتماعی به یک جریان اصلی تبدیل شده است و روابط شخصی را به صورت اصولی تغییر داده است، این رسانه‌ها موضوعات مشارکتی را ایجاد کرده است، امکانات و چالش‌های جدیدی را به وجود آورده است. این موضوع باعث شده است که سازمان علاوه بر توسعه محصول جدید، به اشتراک دانش نیز توجه داشته باشند و توصیه‌هایی به سازمان جهت تشویق کاربران برای استفاده از رسانه‌هایی اجتماعی در راستای مدیریت دانش می‌دهد. در[۳۱] نشان دادند استفاده از رسانه‌های اجتماعی باعث ایجاد صحت دانش می‌شود. دانش در مورد اینکه چه کسی، چه چیزی را می‌داند، و چه کسی، چه کسی را می‌شناسد، اطلاعاتی می‌دهد. این موضوع باعث کاهش چسبندگی دانش می‌شود، سرمایه اجتماعی کارمندان را افزایش می‌دهد و باعث افزایش عملکرد آن‌ها می‌شود.

۳- روش پیشنهادی

در این بخش به ارائه روش پیشنهادی برای تشخیص پست‌های اثرگذار جهت بهبود مدیریت دانش در رسانه‌های اجتماعی پرداخته می‌شود. این روش IKS نامیده می‌شود و فرآیند آن در شکل ۱ نشان داده شده است.



شکل (۱): معماری پیشنهادی برای پیش بینی رفتار بازنشر کاربران مبتنی بر یادگیری عمیق

۳-۱- پیش پردازش داده:

شامل تغییر و تبدیل ویژگی‌ها و نرمال‌سازی می‌باشد، اولین مرحله در مکانیزم پیش‌بینی رفتار بازنشر کاربران مبتنی بر یادگیری عمیق پیشنهادی می‌باشد که اطلاعات مهم را از داده‌های در مقیاس بزرگ در محیط شبکه اجتماعی مورد بررسی قرار داده و انتخاب می‌نماید.

۳-۲ - تبدیل ویژگی:

از آنجایی که مدل پیشنهادی فقط ویژگی‌های عددی را قبول می‌کند، هر مقدار ویژگی سمبولیک به یک مقدار عددی تغییر می‌یابد. به عنوان مثال، مجموعه داده جمع‌آوری شده تعدادی ویژگی سمبولیک همانند منفی بودن توئیت وجود دارد که شامل توئیت منفی، مثبت و یا خنثی می‌باشد که به مقادیر ۲،۱ و ۳ نگاشت می‌شوند.

۳-۳ - اندازه‌گیری ارزش محتوایی یک توئیت

برای ارزش محتوایی در منابع گوناگون تعاریف غیررسمی متعددی ارائه شده است. در این رساله ارزش محتوایی یک توئیت به میزان توجهی که یک توئیت در مخاطب خود ایجاد می‌کند، اطلاق می‌گردد. این توجه در قالب انجام رخداد بازنشر، پاسخ و یا ذکرکردن توسط کاربر نشان داده می‌شود. همان طور که گفته شد روش پیشنهادی برای اندازه‌گیری آن، بر اساس اندازه‌گیری مبتنی بر محتوا و مبتنی بر آمار توئیت می‌باشد، که در ادامه به بررسی آنها پرداخته می‌شود. مواردی که در این رساله جهت اندازه گیری ارزش محتوا مورد سنجش قرار می گیرد در ادامه به صورت جزئی توضیح داده خواهد شد.

۳-۳-۱ - بررسی تازگی یک توییت

در ارتباط با بررسی محتوای یک توئیت از جهت سنجش ارزش محتوایی کارهای بسیاری در زمینه‌های خبررسانی صورت گرفته است. مجموعه خصوصیات‌های فراوانی برای بررسی محتوا ارائه شده است که شامل یک سری از خصوصیات مربوط به پیغام (طول، سوالی بودن یا نبودن و غیره)، کاربر نویسنده (تعداد پیگیری‌کنندگان، سال ثبت نام و غیره)، موضوع پیغام می‌باشند. این دسته از روش‌ها با بررسی لغوی کلمات و تکنیک‌های آماری به دسته بندی، رتبه‌بندی موضوعی و شباهت

$$Novelty = \frac{topic - sim(tw_i, P_{u_i})}{sim(M_{topic\#} - \{M_{topic\#} \cap tw_i\}, P_{u_i})} \quad (۲)$$

سنجی‌های مورد نیاز می‌پردازند. در روش پیشنهادی خصوصیات‌های جدیدی برای اندازه‌گیری برخطِ ارزش اطلاعات بر اساس محتوا بر اساس کارکرد مسئله در نظر گرفته شده است. تازگی یک توئیت یک خصوصیت مهم در ارزش محتوایی آن می‌باشد. بنا به تعریف، یک توئیت tw_i برای کاربر U_i تازگی دارد اگر مرتبط با علایق U_i باشد اما برای او ناشناخته باشد.

$P_{u_i} = \{tw_1, tw_2, \ldots, tw_n\}$ مجموعه توئیت‌های ایجاد شده توسط کاربرU_i می‌باشد. همان طور که در تعریف تازگی یک توئیت[۵] اعلام شد، اولا بایستی مرتبط بودن توئیت را با علایق کاربر پیدا کرد. به منظور تعیین میزان ارتباط یک توئیت با موضوعات مورد علاقه کاربر در اغلب کارهای انجام شده از روش TF-IDF و کسینوس زاویه بین بردار کلمات tw_i و P_{u_i} استفاده می‌گردد. این روش با توجه به متنوع بودن کلمات بکارگرفته شده، دارای دقت پایینی می‌باشد. به همین منظور با استفاده از روش مدل‌سازی موضوع تخصیص پنهان دیریکله می‌توان موضوعات مختلف در متون را دسته‌بندی کرد. هر موضوع شامل مجموعه‌ای از کلمات $M_{topic\#} = \{w_1, w_2, \ldots, w_n\}$ می‌باشد که احتمال وقوع آن‌ها در موضوع مربوطه مشخص شده است. برای افزایش دقت سنجش ارتباط بین توئیت tw_i و مجموعه توئیت‌های ایجاد شده توسط کاربر U_i، $P_{u_i} = \{tw_1, tw_2, \ldots tw_n\}$ ، با استفاده از رابط 1 شباهت موضوعی اندازه‌گیری می‌گردد.

$$topic - sim(tw_i, P_{u_i}) = \frac{topic_{tw_i}\, topic_{P_{u_i}}}{\| topic_{tw_i} \| . \| topic_{P_{u_i}} \|} \qquad (۱)$$

با استفاده از روش مطرح تخصیص پنهان دیریکله می‌توان موضوع مربوط به هر توئیت را در قالب یک موضوع شامل دسته‌ایی از کلمات با احتمال رخداد در این موضوع بدست آورد. در نتیجه با استفاده از این روش موضوع توئیت مورد نظر را پیدا کرده و دسته کلمات مرتبط با این وضوع مشخص می‌گردد.

در صورتی توئیت tw_i ابتدا با یک روش استاندارد کلمات ساکن آن حذف شود و مجموعه $tw_i = \{w_1, w_2, \ldots, w_m\}$، مجموعه کلمات استفاده شده در توئیت tw_i، و مجموعه $M_{topic\#} = \{w_1, w_2, \ldots, w_L\}$، مجموعه کلمات مرتبط با موضوع tw_i باشد، رابطه 2 میزان تازگی توئیت را مشخص می‌نماید.

در رابطه ۲ ، $M_{topic\#} - \{M_{topic\#} \cap tw_i\}$ مجموعه کلمات مرتبط با موضوع می‌باشد که در توئیت وجود ندارد. همان‌طور که در رابطه 2 مشخص است، هر چه کلمات مرتبط با موضوع، شباهت بیشتری با موضوعات مورد بحث توسط کاربر را داشته باشد، احتمال تازگی این توئیت کاهش پیدا می‌کند. تازگی توئیت بنا به تعریف با شباهت آن با توئیت‌های ایجاد شده توسط کاربر نسبت مستقیم دارد.

۳-۳-۲ - *بررسی شباهت جغرافیایی یک توییت*

علاوه بر تازگی توئیت، ارتباط یک توئیت با کاربر می‌تواند نکته کلیدی در ارزش محتوایی این توئیت توسط کاربر باشد. در این رساله علاوه بر محاسبه شباهت لغوی بین توئیت tw_i و توئیت‌های بکار گرفته شده توسط کاربر، شباهت جغرافیایی در نظر گرفته شده است. به طور معمول میزان اثرگذاری توئیت‌هایی که به مباحث منطقه‌ای می‌پردازند، بیش از تویئت‌های سایر مناطق است. به همین جهت در این رساله از مجموعه داده maxmind شامل ۴ میلیون اسامی شهرها و مناطق مختلف به همراه اطلاعات تکمیلی از کشور، موقعیت و غیره برای یافتن کلمات مربوط به شهرها و مناطق جغرافیایی استفاده می‌گردد. برای یافتن کلمات مربوط به شهرها از توئیت tw_i تمام کلمات موجود در توئیت به جز کلمات اضافی مورد استفاده قرار می‌گیرد (حتی هشتگ‌ها با حذف نماد #). درصورتی که LOC_{tw_i} مجموعه شهرها به علاوه کشور و منطقه استفاده شده در توئیت باشند و LOC_{U_i} مجموعه شهرها، کشور و منطقه‌های استفاده شده توسط کاربر U_i در کل توئیت‌های ایجاد شده باشند، رابطه ۳ نشان دهنده میزان ارتباط جغرافیایی بین توئیت tw_i و کاربر U_i می‌باشد[۱۱].

$$\left| \left(Loc_{tw_i} \bigcap_{u_i \in U \,\&\, tw_i \in P} Loc_{u_i} \right) \right| \qquad (۳)$$

۳-۳-۳- *بررسی احساسات یک توییت*

در[۲۰] نشان داده شده است که اخبار و اطلاعات منفی قابلیت انتشار بیشتری نسبت به رخدادهای مثبت دارند. به همین دلیل به منظور تحلیل محتوایی یک توئیت برای اندازه‌گیری ارزش محتوایی، در این رساله منفی بودن توئیت مشخص می‌گردد. در روش بکار گرفته شده برای اندازه‌گیری مقدار منفی بودن توئیت سعی شده است از روش ساده برای اجرا بر خط استفاده گردد. برای این کار از مجموعه داده توئیتر دانشگاه میشیگان ستفاده شده است. در این مجموعه داده 1,578,627 توئیت برچسب‌گذاری شده (برچسب ۱ برای توئیت‌های مثبت و ۰ برای توئیت‌های منفی) گردآوری شده است. با استفاده از ابزار پردازش زبان طبیعی پایتون NLTK و روش naïve bayes، تحلیل احساسات توئیت‌ها صورت گرفته است. برای بدست‌آوردن جدا کننده از ۱/۱۰ داده‌ها به صورت تست و بقیه برای آموزش استفاده شده است. به عنوان پیش‌پردازش دو مرحله نشانه‌گذاری و نرمال‌سازی روی مجموعه داده صورت می‌پذیرد. مخففسازی‌ها و ایکن‌های احساسی (OMG, ASAP,WTF…) در مرحله نشانه‌گذاری به صورت یک نشانه مجزا

درنظرگرفته می‌شود. در مرحله نرمال‌سازی کلمات باحروف بزرگ به حروف کوچک تبدیل می‌شوند (برای مثال I LOVE it!!!!) و تکرار حروف در توئیت‌ها (برای مثال I am happyyyy!!) حذف می‌گردند.

۳-۳-۴- بررسی رشد یک توییت

پارامتر دیگری که مقدار ارزش محتوایی یک توئیت را تعیین می‌نماید، رشد موضوع توئیت است. موضوعات توئیت برگرفته از وقایع خارج از محیط میکروبلاگ می‌توانند بر رفتار بازنشر توئیت تاثیر مستقیمی بگذارد. در این رساله با استفاده از روش LDA که پیشتر بدان اشاره شد، موضوعات مربوط به یک توئیت استخراج می‌گردد. درصد رشد موضوع توئیت در بازه زمانی (۲۴ ساعت) با استفاده از رابطه ۴ محاسبه می‌گردد. در رابطه 4، $M_{topic\#}$ مجموعه کلمات مرتبط با موضوع توئیت tw_i و $\#tw_{w_t}^{t-24}$تعداد توئیت‌های ۲۴ ساعت گذشته است که حاوی کلمات $w_t \in M_{topic\#}$ به ازای تمامی توئیت‌های $\forall tw_i \in tw$ می‌باشند[۲۰].

$$opic_{growth} = \left[\frac{1}{24}\left(\frac{\sum_{w_t \in M_{topic\#}} \#tw_{w_t}^{t} - \sum_{w_t \in M_{topic\#}} \#tw_{w_t}^{t-24}}{\sum_{w_t \in M_{topic\#}} \#tw_{w_t}^{t-24}}\right)\right] \times 100 \quad (۴)$$

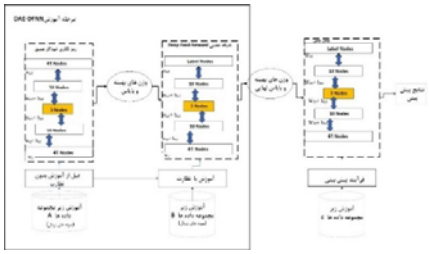
۳-۳-۵- وزن توییت

برای تکمیل تعیین ارزش محتوایی توئیت، از خصیصهٔ مربوط به توئیت نیز استفاده می‌گردد. این خصیصه به صورت مکمل بخش محتوا به گونه‌ایی انتخاب شده است که بتوان به صورت برخط و کارآمد از آن استفاده کرد. خصیصه‌های زیادی برای پیشگویی رفتار بازنشر در مقالات متعددی استفاده شده‌اند، با این حال در انتخاب این خصیصه‌ها دو جنبه قابلیت عملیاتی و مرتبط بودن به ارزش محتوایی لحاظ شده است[۳۳]. با توجه به آمار بازنشرهای یک پست، خصیصه‌ای معرفی می‌گردد که در این رساله وزن یک توئیت گفته می‌شود. وزن توئیت با توجه به اعتبار نویسنده توئیت و کاربرانی که این توئیت را بازنشر کرده‌اند، نسبت مستقیم دارند. اگر $U_{tw_i}^{RT} = \{u_1, u_2, ..., u_R\}$ مجموعه کاربرانی است که تا زمان t، توئیت tw_i را بازنشر کرده‌اند، رابطه ۵ وزن توئیت tw_i را تعیین می‌نماید. در رابطه ۵، $\#f(u_j)$ و $\#E(u_j)$ به ترتیب تعداد کاربران پیگیری‌کننده‌های کاربر U_j و تعداد پیگیری شونده‌ها توسط کاربر U_j می‌باشد. t زمان محاسبه و Au_{tw_i} نویسنده توئیت tw_i را نشان می‌دهد. به علت وجود رابطه پیگیری متقابل، نمی‌توان فقط با استناد به # اعتبار نویسنده یا کاربر را تعیین کرد. هر چه تعداد پیگیری کنندگان کاربر U_j نسبت به تعداد پیگیری شونده‌ها توسط کاربر U_j بیشتر باشد، اعتبار کاربر U_j بیشتر می‌گردد. یکی از جنبه‌هایی مهم و تعیین کننده برای ارزش محتوایی عمر توئیت است. با توجه به مجموعه داده‌های ارزیابی این رساله میزان بازنشر یک توئیت با توجه به مدت زمانی که از ایجاد آن می‌گذرد بسیار شبیه به قانون قدرت می‌باشد. β توان زمان گذشته شده از ایجاد توئیت tw_i می‌باشد که مطابق قانون قدرت به وزن توئیت اعمال شده است. هر چه از زمان ایجاد یک توئیت بگذرد، وزن و ارزش محتوایی آن توئیت کاسته می‌شود.

۳-۴ - یادگیری اثرگذاری با استفاده از روشی مبتنی بر یادگیری عمیق

در این مقاله با استفاده از معماری‌های مختلف مدل یادگیری عمیق برای توسعه یک سیستم پیشگو برای پیش‌بینی رفتار بازنشر کاربران ارائه شده است. در مرحله آموزش، یک الگوریتم رمزنگار خودکار عمیق با استفاده از مشاهدات شبکه‌ای نرمال برای ایجاد پارامترهای اولیه آموزش داده می‌شود. پارامترهای اولیه شامل وزن و بایاس بوده و این الگوریتم یک بازنمایی عمیق از مشاهدات نرمال را می‌آموزد. این پارامترها به عنوان یک مرحله اولیه برای آموزش یک شبکه عصبی پیش خور عمیق استاندارد (IKS) برای یادگیری ضرایب ویژگی‌های موجود و پیش بینی رفتارهای بعدی کاربران در مواجهه با یک پست مورد استفاده قرار می‌گیرد. در مرحله آزمایش، شبکه عصبی پیش خور عمیق استاندارد برای یادگیری ضرایب و تشکیل دسته‌بند مورد استفاده قرار می‌گیرد. گره‌های مختلف پنهان در این تکنیک به صورت حرفه‌ای بازنمایی ویژگی‌های عمیق را یاد گرفته و مهمترین ویژگی‌ها را با تبدیل ابعاد بالای داده به ابعاد پایین مبتنی بر کاهش لایه پنهان دریافت می‌نمایند. جزئیات روش پیشنهادی به تفضیل در ادامه تحقیق شرح داده می‌شود.

ج- فازهای یادگیری و تست سیستم پیش‌بینی رفتار بازنشر مبتنی بر یادگیری عمیق شبکه عصبی پیش خور عمیق (IKS) و خود رمزنگار عمیق (DAE) که در بخش‌های قبلی مورد بحث قرار گرفت، مکانیزم پایه برای ارائه پیش‌بینی رفتار بازنشر کاربران مبتنی بر یادگیری عمیق می‌باشند. ساختار شبکه DAE-DFFFN در شکل 2 نمایش داده شده است.



شکل (۲): معماری پیشنهادی مدل DAE-DFFFN

در فاز یادگیری، مجموعه داده یادگیری نرمال فاقد برچسب A، مجموعه داده یادگیری دارای برچسب B، که در آن ACB، یک خود رمزنگار عمیق (DAE) با یک لایه تنگه تنها با رکوردهای نرمال (A) بدون هر گونه بردار غیر عادی برای یادگیری، آموزش داده می‌شود و مهمترین ویژگی‌ها را برای نماش الگوهای نرمال پوشش می‌دهد. این آموزش با تمام داده‌ها، جایی که ورودی به شبکه (x(i)) از سه لایه پنهان عبور می‌کند، شامل یک تنگه برای بازسازی ((x^(i))، انجام می‌گردد در جایی که We، Wn و Wf وزن‌های شبکه عصبی پیش خور عمیق (IKS)،

خود رمزنگار عمیق (DAE) و مدل پیش‌بینی نهایی هستند که در شکل ۱ توصیف شده است.

در مرحله رمزگذاری، لایه ورودی در اولین لایه پنهان مورد پردازش قرار می‌گیرد. در مرحله تنگه، یک تبدیل غیر خطی با ابعاد پایین از ویژگی ورودی برای استخراج ویژگی‌های موثر در تصمیم بازنشر کاربران اجرا می‌شود. سپس در مرحله رمزگشایی، آخرین لایه پنهان در ویژگی تنگه برای تقریب تکرار ورودی مورد استفاده قرار می‌گیرد. و پس انتشار شیب تصادفی نزولی برای کاهش تابع ضرر یا تابع هزینه یعنی خطای میانگین مربعات مابین (x(i)) و (x^(i)) به مورد استفاده قرار می‌گیرد. فرایند کلیدی برای یادگیری بدون نظارت پیش‌بینی رفتار بازنشر کاربران مبتنی بر یادگیری عمیق پیشنهادی در شکل ۲ ارائه شده است.

سپس مدل آموزشی برای نقطه شروع و مقداردهی اولیه پارامترهای وزن و بایاس برای یادگیری برای شبکه عمیق یادگیری نظارتی و انجام تنظیمات مدل شبکه‌ای با استفاده از مجموعه داده یادگیری برچسب‌دار (B((x(i)), (y(i)))) مورد استفاده قرار می‌گیرد. همان مراحل که قبلا بیان شد، برای یادگیری و اعتبارسنجی روش پیش‌بینی رفتار بازنشر کاربران مبتنی بر یادگیری عمیق پیشنهادی روی مجموعه داده B که شامل داده‌های مربوط به رفتار بازنشر و عدم بازنشر برای تست دقت روش پیشنهادی می‌باشد، نیز دنبال می‌شود. با جزئیات بیشتر، مدل شبکه‌ای مبتنی بر مکانیزم پس انتشار شیب تصادفی نزولی برای به حداقل رساندن تابع ضرر با خطای میانگین مربعات محاسبه شده از تفاوت مابین مقدار خروجی هدف (y(i)) و خروجی پیش‌بینی شده $g\theta'(x(i))$ آموزش داده می‌شود که $g\theta'$ تابع فرضی است که حاصل یک خروجی تخمینی است. در فاز تست، پس از اینکه پارامترها به صورت خودکار در فاز یادگیری، آموزش داده شدند، نمونه مجموعه داده جدید (C ⊆ [A, B]) مبتنی بر مدل شبکه ساخته شده نهایی مورد آزمایش قرار می‌گیرد. هر رکورد ورودی (x^(i)) به لایه ورودی با مقدار اولیه وزنی و بایاس θf = (Wf,bf) تطبیق داده شده و سپس داده ورودی از طریق لایه‌های پنهان مورد پردازش قرار می‌گیرد. سرانجام، لایه خروجی کلاس داده ورودی را به عنوان موارد بازنشر یا عدم بازنشر مبتنی بر مقدار تخمینی تابع ضرر هر کلاس (۵) پیش‌بینی می‌نماید.

۴- ارزیابی

برای ارزیابی چندین سناریو در نظر گرفته شده است. سه مجموعه داده برای ارزیابی میزان تخمین اثرگذاری توئیت‌ها به کار گرفته شده است که با روش‌های مختلف به یادگیری خصیصه‌های موثر در ارزش محتوایی یک توئیت پرداخته شده است. روش دیگر ارزیابی مبتنی بر پرسش‌نامه می‌باشد.

۴-۱ - روش شهودی

روش شهودی به وسیله پرسشنامه که با روش دلفی در سه مرحله انجام شده است. در این روش با در نظر گرفتن تعدادی از کاربران، به صورت روزانه ۵۰-۶۰ توئیت به منظور بررسی برای بازنشر در اختیار آنها قرار گرفته است. کاربران با مطالعه توئیت‌ها در مورد منفی بودن توئیت مرتبط بودن آن و تازگی آن اظهار نظر می‌نمایند. با استفاده از این دو روش می‌توان یافته‌های مطرح شده در این تحقیق ارزیابی کرد. پیش از ارزیابی روش پیشنهادی بایستی روش برآورد خصیصه‌های مطرح شده، ارزیابی گردد. به همین منظور در کنار ارزیابی به کمک مجموعه داده، با استفاده از یک گروه داوطلب از کاربران، به ارزیابی روش پیشنهادی پرداخته می‌شود. در این روش ۴۸ کاربر با سنین بین ۲۰ الی ۴۰ سال انتخاب شده است و طول دوره ارزیابی ۱۰ روز انتخاب شده است. هر کاربر حداقل ۲۰ توئیت ایجاد شده توسط خود را به صورت داوطلبانه در اختیار آزمون قرار می‌دهد. برای کاربران داوطلب به طور روزانه ۵۰-۶۰ توئیت جهت ارزیابی در نظر گرفته شده است. داوطلبان به صورت مستقل به صورت روزانه در این ارزیابی شرکت می‌کنند و نظرات دیگر کاربران اطلاعاتی ندارند. این داوطلبان ضمن بررسی این توئیت‌ها برای بازنشر، به سوالات مشخص شده در ارتباط با هر یک از خصیصه‌های پیشنهادی (منفی بودن، تازگی و ارتباط با علایق) توئیت‌ها پاسخ می‌دهند. سوالات مطرح شده در ارتباط با تازگی پست‌ها و میزان ارتباط توئیت با خود بین ۱تا ۵ امتیازدهی می‌نمایند. برای خصیصه منفی بودن به صورت باینری منفی بودن یک توئیت مورد سوال قرار می‌گیرد. مقدار ۱ نشان‌دهندهٔ منفی بودن این توئیت از دیدکاربر است. در پاسخ‌های ۱ تا ۵، ۱ نشان دهنده کمترین مقدار ارزیابی شده از خصیصه مورد سوال است و ۵ نشان دهنده بیشتر مقدار ارزیابی شده از این خصیصه می‌باشد. برای ارزیابی روش پیشنهادی، در مجموع ۵۳۱ توئیت توسط ۴۸ کاربر داوطلب ارزیابی گردید نتایج این ارزیابی در جدول ۱ قابل مشاهده است. برای مقایسه مقادیر اندازه‌گیری شده برای دو خصیصه (تازگی تویئت و ارتباط توئیت با کاربر) با مقادیر بیان شده توسط کاربران، از ضریب همبستگی کاندال تاو استفاده می‌گردد. این روش برای بدست آوردن همبستگی رنبه بندی‌های ارائه شده دو کمیت بکارگرفته می‌شود. در صورتی که y_i مقدار بیان شده از هر کدام از دو خصیصه (تازگی توئیت و ارتباط) توسط کاربران باشد و x_i مقدار اندازه‌گیری شده پیشنهادی

$$weight(tw_i,t) = \left[\sum_{u_j \in tw_i^t} log(\frac{\#f(u_j)}{\#E(u_j)}) + log(\frac{\#f(Au_{tw_i})}{\#E(Au_{tw_i})})\right] \times (t - t_{tw_i})^{-\beta}$$

باشد و $(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n)$ مقادیر بدست آمده برای کل n توئیت در نظر گرفته می‌شود. برای هر زوج $(x_i,y_i)\&(x_j,y_j)$ به شرطی که $i \neq j$ باشد، هم نوا اطلاق می‌گردد در صورتی که $xi > xj$ و $yi > yj$ باشد یا در صورتی که $xi < xj$ باشد، $yi < yj$. نیز باشد. این دو زوج را غیرهم نوا گوییم در صورتی که اگر $xi > xj$ باشد $yi < yj$ باشد و یا در صورتی که $xi < xj$ باشد، در صورتی تساوی این دو کمیت، نه هم نوا یا غیر هموا اطلاق نمی‌گردد. ضریب همبستگی کاندال بر اساس رابطه (6) تعیین می‌گردد. مقدار ضریب همبستگی کاندال بین ۱ و ۱- می‌باشد ($-1 \leq \tau \leq 1$). هر چه مقدار رتبه یک توئیت بر اساس دو خصیصه، نزدیک به رتبه آن توئیت با توجه

به بیان کاربران (در پرسش نامه) باشد، مقدار این ضریب به ۱ نزدیک‌تر خواهد بود. در صورتی که این رتبه بندی‌ها کاملا متفاوت باشد این ضریب به ۱- نزدیک خواهد بود و مقدار نشان دهنده مستقل بودن این دو مقدار از هم است.

در این قسمت، ارتباط این خصیصه‌های با رفتار بازنشر مورد ارزیابی قرار

$$J(retweet, NegativeTweet) = \frac{M_{11}}{M_{01} + M_{10} + M_{11}} \quad (۷)$$

می‌گیرد. در صورتی که میزان این خصیصه‌ها ارتباط معناداری با رفتار بازنشر داشته باشد می‌توان این ارتباط را با استفاده از یادگیری‌های مرسوم اندازه‌گیری کرد. به همین منظور از روش مشابه با روش همبستگی پیرسون استفاده می‌شود. با توجه به اینکه مقادیر هر یک از این خصیصه‌ها به صورت پیوسته است و رفتار بازنشر به صورت یک مقدار باینری (۰ و ۱) می‌باشد، نمی توان از روش پیرسون استفاده کرد. روش point-biserial برای این گونه مسائل بکارگرفته می‌شود. در صورتی که مقادیر هر یک از خصیصه‌ها را برای هر توئیت به صورت متغیر پیوسته x در نظر گرفته شود و مقدار رفتار بازنشر آن توئیت به صورت باینری y در نظر گرفته شود. ضریب همبستگی point-biserial بر اساس رابطه (7) می‌باشد که در آن M_1 میانگین مقدار خصیصه برای توئیت‌هایی است که منجر به رفتار بازنشر y=1 می‌گردند. به همین صورت M_0 میانگین مقدار خصیصه برای توئیت‌هایی است که منجر به رفتار بازنشر y=1 می‌گردند. n_1 تعداد توئیت‌هایی است که در نمونه‌های بررسی شده جز توئیت‌های منجر به رفتار بازنشر y=1 می‌گردند و n_0 تعداد توئیت‌هایی است که در نمونه‌های بررسی شده جز توئیت‌های منجر به رفتار بازنشر قرار ندارند (y=0). n تعداد کل نمونه‌های بررسی شده است و s_n مقدار انحراف معیار برای مقادیر هر خصیصه‌ها برای کلیه توئیت‌های نمونه بررسی شده می‌باشد. مقدار این ضریب همبستگی بین ۱- تا ۱ می‌باشد که ۱ نشان دهنده بیشترین همبستگی مثبت بین مقادیر اندازه‌گیری شده و رفتار کاربران می‌باشد و ۱- نشان دهنده بیشترین همبستگی منفی بین مقادیر اندازه‌گیری شده و رفتار کاربران است و ۰ نشان دهنده مستقل بودن این خصیصه‌ها با رفتار بازنشر کاربران است.

$$\frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\frac{1}{2}n(n-1)} \quad (۶)$$

به منظور نشان دادن ارتباط خصیصه منفی بودن توئیت با رفتاربازنشر از ضریب شباهت جاکارد رابطه 8 استفاده می‌شود. این ضریب برای نشان‌دادن همبستگی دو متغیر باینری به‌کار گرفته می‌شود. M_{11} مواردی است که مقدار منفی بودن توئیت ۱ (توئیت منفی است) و رفتار بازنشر۱ (رفتار بازنشر اتفاق افتاده است) می‌باشد. در حالت کلی M_{ij} برای دو متغیر باینری منفی بودن توئیت و رفتار بازنشر، که iمربوط به متغیر منفی بودن و j مربوط به متغیر رفتار بازنشر می‌باشد. جمع حالت‌های مطرح برای این دومتغیر ($M_{00}+M_{01}+M_{10}+M_{11}$) برابر با تعداد نمونه‌های مورد بررسی می‌باشد.

نتایج بررسی‌های انجام شده از ضریب همبستگی بین خصیصه‌ها و رفتار بازنشر در جدول ۲ گردآوری شده است. این نتایج نشان دهنده همبستگی مثبت بین این خصیصه‌ها و رفتار بازنشر می‌باشد. بیشترین همبستگی مربوط به وزن توئیت و شباهت موضوع بوده و کمترین همبستگی مربوط به توئیت‌های منفی می‌باشد.

۴-۲- *مدل‌ها و معیارهای ارزیابی*

برای ارزیابی از معیارهای مطرح برای نشان دادن دقت پیشگویی استفاده می‌شود. این معیارها برای مسائل پیشگویی از دسته مسائل کلاس‌بندی باینری در نظر گرفته شده‌اند. که در فرمول های ۸ تا ۱۲ ارائه شده است.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (۸)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \quad (۹)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (۱۰)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (۱۱)$$

$$\text{F-measure} = \frac{2 * (\text{Precision} * \text{recall})}{\text{Precision} + \text{recall}} \quad (۱۲)$$

مثبت صحیح (TP): یک پست بازنشر شده، به درستی به عنوان پست بازنشر شده، تشخیص داده می‌شود.

مثبت کاذب (FP): یک پست بازنشر نشده، به اشتباه به عنوان پست بازنشر شده، تشخیص داده می‌شود.

منفی صحیح (TN): پست بازنشر نشده، به درستی به عنوان پست بازنشر نشده، تشخیص داده می‌شود.

منفی کاذب (FN): پست بازنشر شده، به اشتباه به عنوان پست بازنشر نشده، تشخیص داده می‌شود.

مجموعه دادگان :

جزئیات بیشتری در ارتباط با مجموعه داده و مباحث آماری آن، در

جدول ۳مطرح شده است.

در ادامه بخش ارزیابی، نتایج ارزیابی روش پیشنهادی و مقایسه آن با دیگر روش‌ها ارائه می‌گردد. در این مقاله برای پیش‌بینی تصمیم بازنشر کاربران روش IKS پیشنهاد شده است. نتایج ارزیابی‌ها نشان‌دهنده کیفیت روش‌های پیشنهادی در مقایسه با دیگر روش‌ها می‌باشد.

معیار دقت، یکی از پرکاربردترین معیارها برای ارزیابی روش‌های پیش‌بینی و کلاس‌بندی است. هر چه نسبت میزان پیش‌بینی درست تصمیم بازنشر، به پیش‌بینی‌های نادرست، بیشتر باشد، دقت روش مربوطه بالاتر خواهد بود. جزئیاتی بیشتری از نتایج مقایسه روش‌های پیشنهادی با دیگر روش‌ها از نظر معیار دقت، در جدول (۴)، آورده شده است.

همان طور که در جدول (۴) مشاهده می‌شود، روش IKS در این معیار، بهترین عملکرد را داشته است. بعد از این روش ماشین بردار پشتیبان بهترین دقت را در بین روش‌های دیگر داشته است. یکی از دلایل برتری روش‌های پیشنهادی بر اساس معیار دقت در ماهیت این روش‌ها می‌باشد. در اجتماعات برخط ممکن است کاربران دچار تغییر سلایق شوند. برای نمونه کاربری که قبلا پست‌های سیاسی را بازنشر می‌کرد و علایق سیاسی داشت دچار تغییر رفتار شده و اقدام به بازنشر پست‌ها با موضوعات دیگر نماید. روش پیشنهادی در مقابل این گونه وضعیت‌ها کاملا انعطاف‌پذیر بوده و مدل جدیدی بر اساس داده‌های جدید ساخته می‌شود. این برتری روش پیشنهادی در دقت پیش‌بینی‌های انجام شده بسیار موثر است. از طرف دیگر در روش پیشنهادی، برای هر کاربر بر اساس معیارهای موثر در تصمیم بازنشر یک مدل طبقه‌بندی ساخته می‌شود، در حالی که در روش‌های دیگر به صورت دسته‌ای برای همه کاربران یک مدل واحد استفاده می‌شود. این امکان در روش پیشنهادی باعث برتری این روش بر روش‌های دیگر پیش‌بینی تصمیم بازنشر شده است.

در جدول (۵)، معیار F-measure روش‌های مختلف مقایسه شده است. همان‌طور که پیش‌تر اشاره شد، معیار F-measure بکارگرفته شده در این مقاله، میانگین موزون دقت و بازیابی (حساسیت) می‌باشد (β=۱).همان‌طور که اشاره شد، میزان این معیار، ارتباط مستقیمی به دقت و بازیابی روش‌ها دارد. هر دو روش پیشنهادی، بهترین عملکرد را داشته‌اند. نکته جالب در این جدول، پایین بودن مقدار F measure برای روش درخت تصمیم است. با وجود دقت نسبتاً خوب این روش، با توجه به ضعف شدید این روش در بازیابی، بدترین عملکرد را در این معیار دارد. لجستیک رگرسیون، در نقطه مقابل، علی‌رغم، دقت پایین، در بخش حساسیت، عملکرد مناسبی داشت. در نتیجه مقدار F-measure برای این روش چندان پایین نیست.

۵- نتیجه‌گیری

در این مقاله به یکی از موضوعات بسیار مهم در حوزه مدیریت دانش پرداخته شده است. اگرچه رسانه‌های اجتماعی یکی از بسترهای مهم برای انتشار دانش می‌باشد، اما کاربران در بسیاری مواقع از اطلاعات مفید آن بهره‌مند نمی‌شود. در این مقاله به ارائه روشی جهت بهبود انتشار اطلاعات مطرح شده است. این روش از چندین مرحله تشکیل شده است که شامل جمع آوری داده، پیش پردازش داده، استخراج ویژگی و یادگیری مدل تشکیل شده است. روش یادگیری در این مقاله به یادگیری عمیق معروف است. برای ارزیابی این مدل از دو روش شهودی برای سنجش ویژگی‌های اسخراج شده و همچنین ارزیابی مجموعه دادگان جهت بررسی دقت و F-measure استفاده شده است.

در آینده می‌توان علاوه بر داده‌های متنی از داده‌ها چند رسانه‌ایی از جمله مجموعه داده یوتویب استفاده کرد که می‌توان ویژگی‌های جدیدی به مدل پیشنهادی اضافه کرد. از طرفی شخصیت کاربران نیز در انتشار اطلاعات و بهبود آن می‌تواند بسیار مهم باشد.

منابع و مأخذ

[1] Allcott, H., Gentzkow, M., Chuan., Y.(2019). **Trends in the Diffusion of Misinformation on Social Media**. Research & Politics 6.2.

[2] Araujo, T., Neijens, P., Vliegenthart. R. (2017). **Getting the Word out on Twitter: The Role of Influentials, Information Brokers and Strong Ties in Building Word-of-Mouth for Brands**. International Journal of Advertising 36.3.

[3] Artzi, Y., Pantel, P., Gamon, M. (2012). **Predicting Responses to Microblog Posts**. in NAACL.

[4] Bi, B., Cho, J. (2016). **Modeling a Retweet Network via an Adaptive Bayesian Approach**. in Proceedings of the 25th International Conference on World Wide Web - WWW '16.

[5] Bertani, R. M., Bianchi, R. A., Costa, A. H. R. (2020). **Combining Novelty and Popularity on Personalised Recommendations via User Profile Learning**. Expert Systems with Applications, 146, 113149.

[6] Brady, W. J. (2017). **Emotion Shapes the Diffusion of Moralized Content in Social Networks**. Proceedings of the National Academy of Sciences 114.28 , 7313-7318

[7] Choudhury, M., Harrigan, P. (2014). **CRM to Social CRM: the Integration of New Technologies into Customer Relationship Management**. 22 (2): 149–176, 2014.

[8] Ding, X.J.(2015). **Research on Propagation Model of Public Opinion Topics Based on SCIR in Microblogging**. Comput. Eng, 51, 20–26.

[9] Dong, J., Cheng, Y., Jin, K., Zhang, H., Liu, Q., Jiang, J., Hu, B. (2017). **Effects of Diffusion and Suction Negative Pressure on Coalbed Methane Extraction and a New Measure to Increase the Methane Utilization Rate Fuel**. 197, 70-81.

[10] Feng, L., Hu, Y., Li, B., Stanley, H.E., Havlin, S., Braunstein, L.A.(2015). **Competing for Attention in Social Media under Information Overload Conditions**. PLoS ONE , 10, e0126090.

[11] Forman, C., Goldfarb, A., Greenstein, S. (2005). **Geographic Location and the Diffusion of Internet Technology.** Electronic Commerce Research and Applications, 4(1), 1-13.

[12] Gerald, C.K. (2017). **The Evolutions of Social Media for Organizational Knowledge Management**. Information and organization, http://dx.doi.org/j.Infandorg.01.001.

[13] Hinz, O., Skiera, B., Barrot, C., Becker, J.U. (2011).

Seeding Strategies for Viral Marketing: An Empirical Comparison. Journal of Marketing, vol. 75. pp. 55–71, 2011.

[14] Hock Yeow, Y., Tong-Ming., L. (2017). **Social Trust: Impacts on Social Influential Diffusion**. International Journal of Web Information Systems 13.2.

[15] Hurley, T. A., Green, C. W. (2005). **Knowledge Management and the Nonprofit industry: A Within and between Approach**. Journal of Knowledge Management Practice, 6(1), 1-10.

[16] Hwong, Y.L., Oliver, M. Van Kranendonk, C., Sammut, C., Seroussi, Y.(2017). **What Makes you Tick? The Psychology of Social Media Engagement in Space Science Communication**. Comput. Human Behav.

[17] Iyengar, R., Van den Bulte, C., Valente, T.W. (2011). **Opinion Leadership and Social Contagion in New Product Diffusion**. Mark. Sci.

[18] Jin, H. (2017).**Detection and Characterization of Influential Cross-Lingual Information Diffusion on Social Networks**. Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee.

[19] John, C., Joshua, S.A. (2014). **Epidemiological Modeling of Online Social Network Dynamics**. arXiv , arXiv:1401.4208.

[20] Kim, S. (2011). **The Diffusion of the Internet: Trend and Causes**. Social Science Research, 40(2), 602-613.

[21] Kwahk, K.Y., Park, D.H. (2016). **The Effects of Network Sharing on Knowledge-Sharing Activities and Job Performance in Enterprise Social Media Environments.** Computers in Human Behavior 55 (2016) 826e839.

[22] Lee, J.R.; Chung, C.W.(2014). **A New Correlation-Based Information Diffusion Prediction**. In Proceedings of the 23rd International Conference on World Wide Web Companion, Seoul, Korea, 7–11; pp. 346–351.

[23] Lewis, K., Gonzalez, M., Kaufman, J. (2012). **Social Selection and Peer Influence in an Online Social Network**. Proc. Natl. Acad. Sci., vol. 109, no. 1, pp. 68–72.

[24] Lu, J., Zhao, P., Hoi, S.C.H.(2016). **Online Passive-Aggressive Active learning.** Mach. Learn.

[25] Luukka, P. (2011). **Feature Selection Using Fuzzy Entropy Measures with Similarity Classifier**. Expert Syst. Appl.

[26] Milton, N. R. (2008). **Knowledge Technologies.** (Vol. 3). Polimetrica sas.

[27] Nesi, P., Pantaleo, G., Paoli, I., Zaza, I.(2018). **Assessing the ReTweet Proneness of Tweets: Predictive Models for Retweeting**. Multimed. Tools Appl.

[28] Nisar, T. M., Prabhakar, G., Strakova, L. (2019). **Social Media Information Benefits, Knowledge Management and Smart Organizations**. Journal of Business Research, 94, 264-272.

[29] Petrovic, S., Osborne, M., Lavrenko, V.(2011). **Rt to Win! Predicting Message Propagation in Twitter**. Int. AAAI Conf. Weblogs Soc. Media

[30] Qu, B., Hanjalic, A., Wang, H. (2014). **Heterogeneous Recovery Rates against Sis Epidemics in Directed Networks.** In Proceedings of the International Conference on Network Games, Control and Optimization, Trento, Italia, 29–31.

[31] Riemer, K., Finke, J., Hovorka, D. (2015). **Bridging or Bonding: do Individuals Gain Social Capital from Participation in Enterprise Social Networks?** In Proceedings of the international conference on Information Systems (ICIS2015)

[32] Standing, C., Kiniti, C. (2011). **How can Organization use Wikis for Innovation?**. Techno vision, School of Management, Edith Cowan University,Australia .

[33] Silva, A., Guimarães, S., Meira, J.W., Zaki, M. (2013). **Profile Rank: Finding Relevant Content and Influential Users Based on Information Diffusion**. In Proceedings of the 7th Workshop on Social Network Mining and Analysis (pp. 1-9).

[34] Shin, J. (2018). **The Diffusion of Misinformation on Social Media: Temporal Pattern, Message, and Source**. Computers in Human Behavior 83.

[35] Sun, K., Chen, S. L. (2011). **Risk Reduction via Prototyping in Customized Product Development**. In Interdisciplinary Design: Proceedings of the 21st CIRP Design Conference. Mary Kathryn Thompson.

[36] Trainor, K. J., Andzulis, J., Rapp, A., Agnihotri, R. (2014). **Social Media Technology usage and Customer Relationship Performance: A Capabilities -Based Examination of Social CRM**. J. Bus. Res., 67 (6): 1201–1208

[37] Vermeulen, B. (2007). **Knowledge Based Method for Solving Complexity in Design Problems.** (Doctoral dissertation, TU Delft, Delft University of Technology).

[38] Wang, C. Yang, X.Y., Xu, K., Ma, J.F. (2014). **Seir-Based Model for the Information Spreading.** over SNS. Tien Tzu Hsueh Pao/Acta Electron. Sin, 42, 2325–2330.

[39] Wu, J., Qiang; B.C., Svore, J.C., Krysta, G. M. (2008). **Ranking , Boosting , and Model Adaptation**. Inf. Retr. Boston..

[40] Wymbs, C. (2011). **Digital Marketing: The Time for a New "Academic major" has Arrived**. Journal of Marketing Education, 0273475310392544.

[41] Xu, R., Li, H., Xing, C. (2013). **Research on Information Dissemination Model for Social Networking Services**. Int. J. Comput. Sci, 2, 1–6.

[42] Yang Zhang, Q.Y., Zhiheng, Xu. (2015). **Predicting Popularity of Messages in Twitter Using a Feature-Weighted Model**. Int. J. Adv. Intell.

[43] Young-Kyu, K. (2018). **Influential Users in Social Network Services: The Contingent Value of Connecting User Status and Brokerage**. ACM SIGMIS Database: The DATABASE for Advances in Information Systems 49.1.

[44] Zaman, T.R., Herbrich, R., Stern, D. (2010). **Predicting Information Spreading in Twitter**. Work. Comput. Soc. Sci. Wisdom Crowds NIPS.

[45] Zaman, T., Fox, E.B., Bradlow, E.T. (2014). **A Bayesian Approach for Predicting the Popularity of Tweets**. Ann. Appl. Stat.