

بررسی تشخیص افسردگی در تویتر

سارا ژولا زاده^{۱*}، علی براتی^۲

۱- گروه مهندسی کامپیوتر، واحد دزفول، دانشگاه آزاد اسلامی، دزفول، ایران،

sara.zholazadh@yahoo.com

۲- گروه مهندسی کامپیوتر، واحد دزفول، دانشگاه آزاد اسلامی، دزفول، ایران،

abarati80@yahoo.com

چکیده

امروزه با رشد روزافزون استفاده از شبکه‌های اجتماعی و رشد تعامل بین کاربران، به راحتی می‌توان به نظرات افراد در حوزه‌های مختلف دست یافت. به علاوه استفاده از شبکه‌های اجتماعی در سرتاسر جهان رشد چشمگیری داشته است. در میان این شبکه‌ها تویتر دارای محبوبیت بیشتری است، به همین دلیل در پژوهش‌های مختلف از جمله جهت نظارت بر سلامت روانی کاربران به عنوان منبع اطلاعاتی مورد استفاده قرار می‌گیرد. از آنجا که افسردگی به شایع‌ترین بیماری روحی و روانی در جهان تبدیل شده، در این مقاله به بررسی تشخیص بیماری افسردگی پرداخته شده است. توییت‌های افراد حاوی احساسات درونی و رفتار آن‌ها است؛ پس می‌توان با تجزیه و تحلیل این داده‌ها به تشخیص بیماری افسردگی پرداخت. اصلی‌ترین گام‌های این فرآیند عبارتند از: جمع آوری داده، یافتن بخش‌های مربوط، یافتن احساسات کلی، کمی کردن احساسات و یکپارچه‌سازی تمام الگوهای به دست آمده برای ایجاد یک الگوی احساسی کلی. تحلیل احساسات در متون نوشتاری یکی از زیرمجموعه‌های متن کاوی است، بنابراین در این مقاله نیز به روش‌های متن کاوی پرداخته شده است.

کلمات کلیدی: آنالیز احساسات، تویتر، یادگیری ماشین، تشخیص افسردگی

۱- مقدمه

افسردگی یکی از شایع‌ترین اختلالات روانی است که درصد قابل توجهی از افراد در دوره‌ای از زندگی خود بدان مبتلا می‌شوند. در سراسر جهان تنها بخش کوچکی از مردم به درمان و یا راه‌های تشخیصی افسردگی دسترسی دارند. برای درمان افسردگی ابتدا باید آن را در افراد تشخیص داد. طبق گزارش سازمان بهداشت جهانی اکثر افرادی که دچار افسردگی هستند، حتی پس از مراجعه به پزشکان بیماری افسردگی آنها اغلب ناشناخته و پنهان باقی می‌ماند در نتیجه بسیاری از این افراد درمان نمی‌شوند [۱]. امروزه استفاده از شبکه‌های اجتماعی در سرتاسر جهان رشد چشمگیری داشته است. شبکه‌های اجتماعی این امکان را فراهم می‌آورد تا کاربران بتوانند علاقه‌مندی‌ها، افکار و فعالیت‌های خودشان را با دیگران به اشتراک بگذارند و دیگران هم در این افکار و فعالیت‌ها با آنان سهیم شوند. تویتر یکی از پرطرفدارترین شبکه‌های اجتماعی در جهان است که به کاربران امکان می‌دهد توییت‌هایی تا حداکثر ۱۴۰ کاراکتر را ارسال کنند. این توییت‌ها می‌توانند حاوی احساسات درونی، نظرات و علاقه‌مندی‌های افراد در زمینه‌های مختلف باشد [۲]. با رشد روزافزون اطلاعات متنی تولید شده توسط کاربران در اینترنت، تجزیه و تحلیل احساسات در متون، زمینه کاری جذابی در بین محققان علوم داده کاوی و پردازش زبان طبیعی شده است. تجزیه و تحلیل یا کاوش احساسات، مطالعه محاسباتی احساسات، نظرات، گرایش‌ها و تمایلات کاربران بر روی موضوعات، اشیاء، خصوصیات و ویژگی‌های آن‌ها در اسناد متنی می‌باشد. در این مقاله به مروری بر کارهای انجام شده در این زمینه پرداخته شده است.

۲- آنالیز احساسات

به طور کلی می توان آنالیز احساسات را در سه گروه تقسیم کرد: روش های مبتنی بر واژگان^۳، روش های مبتنی بر یادگیری ماشین^۴ و روش های ترکیبی. روش های مبتنی بر واژگان متکی بر دیکشنری احساسات است که این دیکشنری، مجموعه ای از واژه ها و کلمات شناخته شده و از پیش تعیین شده احساسی می باشد. این روش به ساختار جملات می پردازد و با کمک یک لیست از کلمات و عبارات احساسی و یک مجموعه قوانین، جهت گیری جملات را مشخص می نماید. این روش به دو گروه مبتنی بر دیکشنری و مبتنی بر پیکره^۵ تقسیم می شود. روش مبتنی بر یادگیری ماشین، الگوریتم های یادگیری ماشین را بکار می برد و از ویژگی های زبانی استفاده می کند. در این روش نیازی به شناخت ساختار جملات وجود ندارد. روش ترکیبی، ترکیبی از هر دو روش قبلی می باشد و روشی متداول است [۳-۵].

۲-۱- روش های مبتنی بر واژگان

در [۶]، یک روش ترکیبی از روش های انتخاب ویژگی و استخراج ویژگی ارائه شده است. برای انتخاب ویژگی از ادغام روش های فراوانی سند، واریانس واژه، میانگین مد، میانگین اختلاف مطلق، آنتروپی و بهره اطلاعات استفاده می کند. ویژگی های بدست آمده از این روش ها به وسیله اجتماع، اشتراک و اجتماع اصلاح شده، ترکیب می شوند و سپس زیرمجموعه نهایی توسط الگوریتم k-means خوشه بندی می شود. امتیاز احساسی هر خوشه با استفاده از دیکشنری احساسات SentiWordNet محاسبه می گردد. این روش بر روی مجموعه داده معیار روبرتز-۲۱۵۷۸ و کلاسیک ۴ اجرا می شود. با مقایسه معیارهای ارزیابی فراخوانی^۷، F-Measure، دقت و صحت، نشان می دهد که این روش عملکرد بهتری نسبت به روش های رقابتی دیگر دارد. سیف^۱ و همکاران روشی به نام SentiCircles را پیشنهاد می دهند که روشی مبتنی بر دیکشنری است. این روش برای تشخیص احساسات هر دو سطح موجودیت و توییت استفاده می شود. برای تولید، ابتدا بردار واژه- مفهوم درون یک دایره هندسی دوبعدی تبدیل می شود که این دایره از نقاط مشخص شده واژه های مفهومی تشکیل می گردد. هر واژه مفهومی در دایره بر اساس زاویه آن (تعریف شده توسط احساسات قبلی) و شعاع آن (تعریف شده توسط درجه همبستگی آن با واژه اصلی) قرار می گیرد. در این مقاله از سه مجموعه داده OMD، HCR و STS-Gold استفاده شده است. مجموعه داده OMD، شامل ۳۲۳۸ توییت است که در اولین مناظره تلویزیونی ریاست جمهوری ایالات متحده در ماه سپتامبر ۲۰۰۸ ساخته شده است که تنها از توییت های مثبت و منفی آن استفاده می شود (۱۰۸۱ توییت با ۳۹۳ توییت مثبت و ۶۸۸ توییت منفی). مجموعه HCR که توسط crawling در مارس ۲۰۱۰ ساخته شد که شامل توییت هایی با هشتم HCR است و با حذف توییت های خنثی، مجموعه داده ۱۳۵۴ توییت با ۳۹۷ توییت مثبت و ۹۵۷ توییت منفی بدست آمد. STS-Gold که یک زیرمجموعه از STS است که شامل ۲۰۳۴ توییت (۱۴۰۲ منفی و ۶۳۲ مثبت) و ۵۸ موجودیت می باشد که بوسیله سه ارزیاب انسانی مختلف به صورت دستی نوشته شده است. در این روش با استفاده از سه دیکشنری احساسات، امتیازدهی و تشخیص احساس قبلی کلمات صورت می پذیرد: ۱- دیکشنری Sentiwordnet، ۲- دیکشنری MPQA subjectivity، ۳- دیکشنری TheWall. این روش با روش SentiStrength و روش های پایه مثل ماشین بردار پشتیبان و حداکثر آنتروپی مقایسه می شود. نتایج نشان می دهد که این روش در صحت و F-Measure برای تشخیص های سطح موجودیت بطور قابل توجهی بهتر از روش های پایه اجرا می شود و در تشخیص احساسات سطح توییت، در دو مجموعه داده بهتر از تکنولوژی های جدید SentiStrength با ۴-۵٪ اختلاف در صحت اجرا می گردد، اما در مجموعه داده سوم F-Measure، ۱٪ کاهش می یابد [۷].

۲-۲- روش های مبتنی بر یادگیری ماشین

¹ Saif

در [۸]، طبقه‌بندی احساسات با استفاده از روش‌های گروهی bagging و boosting همراه با روش‌های انتخاب ویژگی آماری مانند PMI و مربع کای صورت گرفت و مقایسه‌ای بین طبقه‌بندی گروهی و روش‌های پایه و همچنین بین دو روش PMI و مربع کای انجام پذیرفت. در ارزیابی صورت گرفته بر روی مجموعه داده SemEval 2013 (شامل ۹۶۳۵ توییت برای آموزش و ۳۰۰۵ توییت برای تست) روش مربع کای نتایج بهتری از PMI داشت؛ بنابراین از روش مربع کای برای انتخاب ویژگی استفاده شد. برای ارزیابی از ارزیابی متقابل^۱ استفاده می‌گردد. مقایسه‌ها، بهبود صحت و اثربخشی بیشتر روش‌های گروهی را نسبت به روش‌های پایه (ماشین بردار پشتیبان، بیز ساده و حداکثر آنتروپی) نشان می‌دهد.

کاتز^۲ و همکاران، روش ConSent را ارائه کردند که برای آنالیز احساسات هر دو متون منظم و نامنظم (با درجه بالایی از نویز) مؤثر می‌باشد. این روش شامل دو مرحله یادگیری و تشخیص است. فرآیند یادگیری سه مرحله دارد. در مرحله اول واژه‌هایی شناسایی می‌شوند که وجود آن‌ها در متن نشان‌دهنده احساسات مثبت یا منفی است (واژه‌های کلیدی). در مرحله دوم مجموعه‌ای از اصطلاحات مفهومی را برای هر واژه کلیدی شناسایی می‌کنند که این اصطلاحات برای مدل کردن متن در مجاورت کلمات کلیدی ایجاد می‌شوند. در مرحله سوم، این اصطلاحات برای ایجاد ویژگی‌ها در روش تحت نظارت استفاده می‌شوند. یکی از مزایای این روش کاهش تعداد ویژگی‌ها است. برای کاهش ویژگی از الحاق فراداده اسناد و آنالیز چندین قسمت متن به طور هم‌زمان استفاده می‌کنند. در مرحله تشخیص، از روش Rotation Forest برای اختصاص نمره به اسناد جدید استفاده می‌کنند. این روش بر روی سه مجموعه داده بررسی هتل از وب‌سایت مشاورین، بررسی فیلم و توییت‌های آمیخته شد. همچنین از دو مجموعه داده بطور خودکار رونویسی شده‌ی دو مرکز تلفن برای آزمایش استفاده گردید. معیارهای ارزیابی این روش فراخوانی، ROC، F-Measure، دقت و صحت می‌باشد که در مقایسه با روش‌های ماشین بردار پشتیبان، بیز ساده و -Delta TFIDF بعد از ماشین بردار پشتیبان بالاترین صحت را بدست می‌آورد [۹].

در [۱۰]، آنالیز احساسات بر روی داده‌های توییت انجام می‌شود که از مراحل جمع‌آوری داده‌ها توسط، پیش‌پردازش و درنهایت استفاده از روش‌های یادگیری ماشین بیز ساده، حداکثر آنتروپی و ماشین بردار پشتیبان تشکیل می‌شود. نتیجه به دست آمده نشان می‌دهد که ماشین بردار پشتیبان بهترین عملکرد را با ۸۳٫۵۶ درصد و ۸۰٫۱۱ درصد صحت برای واژه‌های کلیدی فیس‌بوک و آیفون دارد و نیز طبقه‌بند بیز ساده از روش حداکثر آنتروپی صحت بیشتری را به دست می‌آورد.

در [۱۱]، روشی با استفاده از ساختار نمودارهای ترکیبی واژه‌ها ارائه شد. این روش از ماژول‌های ساخت گراف ترکیبی واژه‌ها، احساسات کاوی، DocumentGrabber و طبقه‌بندی احساسات سند استفاده می‌کند که باعث بهبود قدرت معرفی کلمات جدید و روابط و حذف نمونه‌های متناقض می‌شود. مجموعه داده استفاده شده در این مقاله دیتاست نقد فیلم است که شامل ۱۰۰۰ نقد مثبت و ۱۰۰۰ نقد منفی می‌باشد. این روش با روش‌های ماشین بردار پشتیبان و بیز ساده مقایسه شده و نتایج حاکی از بهبود صحت را به همراه دارد.

در [۱۲]، به‌وسیله روش‌های فراوانی واژه، TF-IDF و استخراج کلمات نظرگرا ویژگی‌ها انتخاب می‌شوند و ارتباط ویژگی‌ها بر اساس شاخص جینی محاسبه می‌گردد. این روش از ماشین بردار پشتیبان برای طبقه‌بندی استفاده می‌کند. مجموعه داده استفاده شده در این مقاله دیتاست IMDB می‌باشد. مقایسه با سایر روش‌های انتخاب ویژگی در نقد فیلم‌ها نشان دادند که طبقه‌بندی با استفاده از این روش، صحت را بهبود می‌دهد.

در [۱۳]، یک سیستم طبقه‌بندی خودکار پیام‌های توییت در مجموعه دسته‌ی از پیش تعریف شده، ارائه می‌شود. ویژگی‌های استفاده‌شده در این سیستم شامل کلمات در توییت، کلمات از آدرس‌ها نوشته‌شده، کلمات از پروفایل کاربر ارجاع شده، کلمات از نزدیک‌ترین مقاله ویکی‌پدیا می‌باشد. اگر یک توییت شامل آدرس‌ها و کاربران ارجاع شده نباشد، اما در ویکی‌پدیا برای آن صفحه‌ای پیدا شود، فقط توییت و واژگان ویکی‌پدیا استفاده خواهند شد. این روش با استفاده از ترکیب‌های مختلف مجموعه ویژگی‌ها در گروه‌های زیر ارزیابی شده است. ویژگی‌های گروه (FGA) A، گروه (FGB) B و گروه (FGC) C که به ترتیب

¹ cross validation

² Katz

سومین همایش ملی مهندسی کامپیوتر، داده کاوی و داده های حجیم

شامل کلمات توییت، کلمات توییت + پروفایل کاربر + کلمات URL و کلمات توییت + پروفایل کاربر + URL + کلمات و یکی پدیا می باشند. برای طبقه بندی از روش بیز ساده استفاده می شود. معیارهای ارزیابی این مقاله صحت، دقت، فراخوانی و F-Measure است که در مجموعه داده ای شامل ۱۵۰ توییت در هر دسته (از قبیل ورزشی، سیاسی و غیره) و در مجموع ۱۰۵۰ توییت مورد بررسی قرار می گیرد. با توجه به نتایج بدست آمده، ترکیبی از مجموعه ترکیبی ویژگی های کلمات توییت با پروفایل کاربر و کلمات استخراج شده از آدرس های مرتبط به بالاترین صحت ۸۰٫۸٪ می رسند و کلمات صفحات و یکی پدیا، کاهش صحت را نشان می دهد. بالاترین نتایج در سایر معیارهای مورد بررسی در گروه FGB و با حدود ۹۰٪ تا ۹۱٪ مشاهده می شود.

۲-۳- روش های ترکیبی

در [۱۴]، یک روش ترکیبی به نام Tinga برای استخراج ویژگی ها از جملات اساسی اسپانیایی ارائه می شود. همچنین در این مقاله، دیکشنری نیمه خودکار از احساسات اسپانیایی بوجود می آید. در پیش پردازش به جملات بر این اساس برچسب می زند: Terrible(N+), bad (N), neutral (NEU), good (P) and excellent (P+) این روش، گروه های احساسی را بر اساس چند قانون تعریف شده به عنوان ویژگی معرفی می کند. این ویژگی ها با روش ماشین بردار پشتیبان طبقه بندی می شوند. در این مقاله از دو مجموعه داده TASS 2014 و مجموعه نقد SFU استفاده شده است. نتایج ارزیابی صحت ۸۷٫۹٪ در 5-fold cross validation، ۵۲٫۶٪ صحت برای مجموعه TASS 2014 و ۶۵٫۰۴٪ برای مجموعه نقد SFU را نشان دادند.

در [۱۵]، یک روش مبتنی بر طبقه بندی چند برچسبی ارائه می شود. ویژگی های مورد استفاده در این مقاله شامل کلمات اولیه قطعه قطعه شده و ویژگی های احساسات می باشد. ویژگی های احساسات بر اساس سه دیکشنری احساسات مختلف، DUTSD، NTUSD و HD و کیسه ای از کلمات است که این کیسه کلمات برای نمایش ویژگی استفاده می شود. در اینجا برای تقسیم بندی متن از یک بسته گایتون استفاده شده است. برای بدست آوردن کلمات ناشناخته در jieba، یک روش بر اساس مدل مخفی مارکوف استفاده شده و با الگوریتم ویتربی حل می شود. در مجموع ۱۱ روش طبقه بندی چند برچسبی BR، CC، CLR، هومر، RAKEL، ECC، MLKNN، RF-PCT، BRKNN-A، BRKNN-FOF و B بر اساس دو مجموعه داده میکرو بلاگ مقایسه می شود. مجموعه داده استفاده شده در این مقاله، HR و IA هستند. ۸ معیار ارزیابی استفاده شده در این مقاله Hamming loss، subset، صحت، F1، micro F1، macro F1، میانگین دقت، پوشش و one error هستند. آزمایش با کلمات قطعه قطعه خام نشان می دهد که RAKEL و ECC به وضوح بهترین هستند و بهترین روش در HR به ترتیب CLR و هومر می باشند؛ همچنین ECC بهترین روش در IA است. در مقابل، BRkNN بدترین روش در هر دو IA و HR می باشد. نتایج کلیدی متعددی در آزمایش با ویژگی های مختلف مشاهده است. به عنوان مثال، DUTSD بهترین عملکرد طبقه بندی چند برچسبی را در میان سه دیکشنری احساسات متفاوت دارد. در HR، ECC بطور کلی بهتر است در حالی که MLkNN و BRkNN بدترین روش ها هستند. در IA، CLR و هومر بهتر از بقیه روش ها برای بیشتر معیارها هستند. با این حال، اجرای ضعیفی در صحت زیرمجموعه ها دارد.

در [۱۶]، برای انتخاب ویژگی از بهره اطلاعات، بایگرم و روش های استخراج شیء گرا استفاده شد. در اینجا پس از انتخاب ویژگی، استاندارد سازی ویژگی ها درون یک دیکشنری صورت می گیرد. در مرحله بعد یک طبقه بند بیز ساده برای دسته بندی کردن توییت های سبجکتیو^{۱۲} و توییت های آبجکتیو^{۱۳} ساخته می شود. برای طبقه بندی توییت های سبجکتیو درون کلاس مثبت یا منفی از ماشین بردار پشتیبان استفاده می گردد. مجموعه داده استفاده شده، ۲۰۰۰۰۰ توییت احساسی از دانشگاه

استنفورد و ۲۰۰۰۰ توییت ساجکتیو از راویکیران جاناراهانا^۱ است. این روش صحت مناسبی در داده‌های آزمایشی خود به دست می‌آورد.

۳- تشخیص افسردگی

در سال ۲۰۱۳ برای نخستین بار ریسینگ^۲ و همکاران، جهت تشخیص افسردگی در اشخاص به تجزیه و تحلیل داده‌های متنی نوشته شده توسط آن‌ها پرداختند و این داده‌ها را فرموله کردند. آن‌ها موضوعاتی از دست نوشته‌های افراد استخراج کرده و با استفاده از LDA توانستند به مدل‌سازی موضوع در یادگیری ماشین دست یابند. با استفاده از مدل‌سازی موضوعات بدست آمده و روش‌های آماری آن‌ها توانستند بیماری افسردگی افراد را تخمین بزنند [۱۷].

مورنو^۳ و همکاران در سال ۲۰۱۱ به بررسی علائم افسردگی دانشجویان در شبکه اجتماعی فیس بوک پرداختند و با بررسی کامنت‌های آن‌ها در این شبکه اجتماعی به میزان افسردگی افراد پی بردند [۱۸].

پارک^۴ و همکاران در سال ۲۰۱۳ به تجزیه و تحلیل فعالیت‌ها و توییت‌های بین کاربران بدون افسردگی و کاربران افسرده پرداختند. تجزیه و تحلیل متن‌های توییت آن‌ها مشابه با تجزیه و تحلیل داده‌ها انجام شده از فیسبوک بود [۱۹].

بولن^۵ و همکاران در سال ۲۰۱۱ توانستند از طریق آنالیز احساسات توییت‌ها خلق و خوی مختلف را استخراج کنند از جمله تنش، افسردگی، خشم، قدرت، خستگی و گیجی دست یابند [۲۰].

دی چودهری^۶ و همکاران در سال ۲۰۱۳ به جمع‌آوری اطلاعات مختلف رفتاری افراد انگلیسی زبان پرداختند. آن‌ها افرادی با ویژگی‌های مختلف را به کمک ماشین بردار پشتیبان دسته‌بندی کردند و از طریق پیشینه آن‌ها توانستند احتمال افسردگی را پیش‌بینی کنند، علاوه بر آن دریافتند که که افسردگی افراد با کاهش میزان استفاده از شبکه‌های اجتماعی ارتباط دارد [۲۱].

در روش‌های بیان شده برای تحلیل احساسات افراد و نوشته‌های آن‌ها از پزشکان و یا مشاوران کمک گرفته شده است، امروزه با توجه به حجم زیاد داده‌ها در شبکه‌های اجتماعی استفاده از امتیازبندی انسانی هزینه‌بر و وقت‌گیر است. ریسینگ^۷ و همکاران در سال ۲۰۱۵ با استفاده از مدل‌سازی موضوع و آنالیز احساسات با استفاده از دیکشنری LIWC توانستند به پیش‌بینی افسردگی دست یابند [۲۲].

تی سوگاوا^۸ و همکاران در سال ۲۰۱۵ با استفاده از مدل رگرسیون چندگانه با تقریب ۰٫۴۵ توانستند به پیش‌بینی و تشخیص افسردگی افراد دست یابند [۲۳].

نادم^۹ و همکاران در سال ۲۰۱۶ از روش Bag of Words جهت کمی کردن توییت‌ها استفاده کردند، آن‌ها به تخمین و پیش‌بینی افسردگی حاد پرداختند. همچنین آن‌ها از چندین طبقه‌بند برای ارزیابی خطر افسردگی استفاده کردند. با استفاده از مجموعه داده‌ای شامل دو میلیون و نیم توییت توانستند به بهترین دقت ۸۶ درصد دست یابند [۲۴].

مورئاس^{۱۰} و همکاران در سال ۲۰۱۳ برای دسته‌بندی توییت‌های خود از الگوریتم SVM استفاده کردند و توانستند دقت دسته‌بندی را تا ۴۵٪ نسبت به روش‌های قبل افزایش دهند [۲۵].

¹ Ravikiran Janardhana

² Resnik

³ Moreno

⁴ Park

⁵ Bollen

⁶ De Choudhury

⁷ Resink

⁸ Tsugawa

⁹ Nadeem

¹⁰ Moraes

شارت^۱ و همکاران در سال ۲۰۱۷ مطالعاتی با هدف پیش‌بینی بیماری روحی با استفاده از رسانه‌های اجتماعی انجام دادند. کاربران مبتلا به بیماری‌های روانی با استفاده از نظرسنجی، اشتراک‌گذاری‌های عمومی آن‌ها در توییتر یا عضویت کاربران در انجمن‌های آنلاین شناسایی شده و آن‌ها از طریق الگوها در زیان و فعالیت آنلاین آن‌ها قابل تشخیص هستند [۲۶].

تائو^۲ و همکاران ۲۰۱۶، در مقاله‌ای با عنوان "تحلیل احساسات برای تشخیص افسردگی در شبکه‌های اجتماعی"، تکنیک‌های تحلیل احساسات را برای تحلیل مشارکت کاربران شبکه اجتماعی در تشخیص افسردگی بصورت بالقوه به کار گرفته‌اند. تشخیص افسردگی از طریق دانش پایه افسردگی و الگوریتمی برای تجزیه و تحلیل داده‌های متنی انجام می‌گیرد [۲۷].

در [۲۸]، محققین تکنیکی برای طبقه‌بندی داده‌های دانشجویان در توییتر، در دسته‌های مختلف در برخورد دانشجویان با مسائل مختلف، پیشنهاد دادند. در [۲۹]، محققین رویکردی منطقی برای کاوش عواطف به اشتراک گذاشته در پلت‌فرم‌های مختلف رسانه اجتماعی، معرفی کردند. در این کار، احساسات متن با استفاده از گرامر دسته‌ای ترکیبی، نشانه‌گذاری، دریافت لغوی و شبکه‌های معنایی، تحلیل می‌شود. تکنیک‌های پایه طبقه‌بندی عاطفی و روش‌های جمع‌آوری داده، در [۳۰] معرفی شده است. دقت فرآیند طبقه‌بندی با بردار ویژگی انتخابی، برای محصولات الکترونیکی با استفاده از طبقه‌بندی کننده‌های مختلف مانند بیز ساده، حداکثر آنتروپی، ماشین بردار پشتیبان و طبقه‌بندی کننده‌های دسته‌ای در [۳۱] تصدیق شد.

۴- نتیجه‌گیری

در دنیای امروز حجم عظیمی از اطلاعات به صورت متن می‌باشد. بنابراین تکنیک‌های متن‌کاوی اهمیت یافته‌اند. کاوش نظرات یا تحلیل احساسات به عنوان شاخه‌ای از متن‌کاوی، به معنی یافتن دیدگاه نویسنده متن درباره یک موضوع خاص است. اینترنت این امکان را برای کاربران فراهم می‌کند تا نظرات خود را به سهولت بیان کنند و از نظرات دیگران در مورد موضوعی خاص مطلع شوند. حجم بالا و فقدان ساختار مناسب برای متن نظرات ارائه شده بروی بستر وب، استفاده از دانش پنهان درون آن‌ها را دشوار نموده است. بنابراین ارائه روش‌هایی که بتواند این دانش را به صورت خلاصه و ساخت یافته آماده کرده و در اختیار ما قرار دهد حائز اهمیت است. در این مقاله، به مروری بر مهم‌ترین و اخیرترین تحقیقات در زمینه تشخیص بیماری افسردگی بر اساس توئیت‌های کاربران در رسانه اجتماعی توییتر پرداخته شد.

۱۰- مراجع

- [1]. Pratt, L. A., & Brody, D. J. (2015). Depression in the US household population, 2009–2012. 2015. Retrieved from Washington, DC: <https://www.cdc.gov/nchs/data/databriefs/db172.pdf>.
- [2]. Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., & Bierut, L. J. (2016). A content analysis of depression-related tweets. *Computers in human behavior*, 54, 351-357.
- [3]. Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14-46.
- [4]. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113.
- [5]. Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data* (pp. 415-463). Springer, Boston, MA.
- [6]. Nebu, C. M., & Joseph, S. (2016). A hybrid dimension reduction technique for document clustering. In *Innovations in bio-inspired computing and applications* (pp. 403-416). Springer, Cham.
- [7]. Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing & Management*, 52(1), 5-19.

¹ sharat

² Tao

- [8]. Devi, K. L., Subathra, P., & Kumar, P. N. (2015). Tweet sentiment classification using an ensemble of machine learning supervised classifiers employing statistical feature selection methods. In Proceedings of the Fifth International Conference on Fuzzy and Neuro Computing (FANCCO-2015)(pp. 1-13). Springer, Cham.
- [9]. Katz, G., Ofek, N., & Shapira, B. (2015). ConSent: Context-based sentiment analysis. Knowledge-Based Systems, 84, 162-178.
- [10]. Chinthala, S., Mande, R., Manne, S., & Vemuri, S. (2015). Sentiment analysis on twitter streaming data. In Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1(pp. 161-168). Springer, Cham.
- [11]. Colace, F., Casaburi, L., De Santo, M., & Greco, L. (2015). Sentiment detection in social networks and in collaborative learning environments. Computers in Human Behavior, 51, 1061-1067.
- [12]. Manek, A. S., Shenoy, P. D., Mohan, M. C., & Venugopal, K. R. (2017). Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. World wide web, 20(2), 135-154.
- [13]. Theodotou, A., & Stassopoulou, A. (2015, November). A system for automatic classification of twitter messages into categories. In International and Interdisciplinary Conference on Modeling and Using Context (pp. 532-537). Springer, Cham.
- [14]. Gutiérrez, E., Cervantes, O., Báez-López, D., & Sánchez, J. A. (2015, June). Sentiment groups as features of a classification model using a spanish sentiment lexicon: A hybrid approach. In Mexican Conference on Pattern Recognition (pp. 258-268). Springer, Cham.
- [15]. Liu, S. M., & Chen, J. H. (2015). A multi-label classification based approach for sentiment classification. Expert Systems with Applications, 42(3), 1083-1093.
- [16]. Le, B., & Nguyen, H. (2015). Twitter sentiment analysis using machine learning techniques. In Advanced Computational Methods for Knowledge Engineering (pp. 279-289). Springer, Cham.
- [17]. Resnik, P., Garron, A., & Resnik, R. (2013). Using topic modeling to improve prediction of neuroticism and depression in college students. In Proceedings of the 2013 conference on empirical methods in natural language processing (pp. 1348-1353).
- [18]. Moreno, M. A., Jelenchick, L. A., Egan, K. G., Cox, E., Young, H., Gannon, K. E., & Becker, T. (2011). Feeling bad on Facebook: Depression disclosures by college students on a social networking site. Depression and anxiety, 28(6), 447-455.
- [19]. Park, S., Lee, S. W., Kwak, J., Cha, M., & Jeong, B. (2013). Activities on Facebook reveal the depressive state of users. Journal of medical Internet research, 15(10).
- [20]. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of computational science, 2(1), 1-8.
- [21]. De Choudhury, M., Counts, S., & Horvitz, E. (2013, April). Predicting postpartum changes in emotion and behavior via social media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 3267-3276). ACM.
- [22]. Resnik, P., Armstrong, W., Claudino, L., Nguyen, T., Nguyen, V. A., & Boyd-Graber, J. (2015). Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (pp. 99-107).
- [23]. Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., & Ohsaki, H. (2015, April). Recognizing depression from twitter activity. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 3187-3196). ACM.
- [24]. Nadeem, M. (2016). Identifying depression on Twitter. arXiv preprint arXiv:1607.07384.
- [25]. Moraes, R., Valiati, J. F., & Neto, W. P. G. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. Expert Systems with Applications, 40(2), 621-633.

- [26]. Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18, 43-49.
- [27]. Tao, X., Zhou, X., Zhang, J., & Yong, J. (2016, December). Sentiment analysis for depression detection on social networks. In *International Conference on Advanced Data Mining and Applications* (pp. 807-810). Springer, Cham.
- [28]. Chen, X., Vorvoreanu, M., & Madhavan, K. (2014). Mining social media data for understanding students' learning experiences. *IEEE Transactions on Learning Technologies*, 7(3), 246-259.
- [29] Kasture, N. R., & Bhilare, P. B. (2015, February). An Approach for Sentiment analysis on social networking sites. In *Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on* (pp. 390-395). IEEE.
- [30] Bhuta, S., Doshi, A., Doshi, U., & Narvekar, M. (2014, February). A review of techniques for sentiment analysis Of Twitter data. In *Issues and challenges in intelligent computing techniques (ICICT), 2014 international conference on* (pp. 583-591). IEEE.
- [31] Neethu, M. S., & Rajasree, R. (2013, July). Sentiment analysis in twitter using machine learning techniques. In *Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on* (pp. 1-5). IEEE.