

مدیریت احساسات مشتریان در رسانه های اجتماعی جهت بهبود تبلیغات و افزایش خرید

لیلا خواجهوند¹، عباس طلوعی اشلقی^{2*}، مرتضی موسی خانی

¹ دانشجوی کارشناسی ارشد، گروه مدیریت فناوری اطلاعات، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

² استاد، گروه مدیریت صنعتی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران (عهده دار مکاتبات)

³ استاد، گروه مدیریت دولتی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

تاریخ دریافت: اسفند 1401، اصلاحیه: فروردین 1402، پذیرش: تیر 1402

چکیده:

امروزه شبکه های اجتماعی توجه ویژه ای را به خود جلب نموده اند. در شبکه های اجتماعی گوناگون، کاربران دائماً در حال ابراز نظرات عمومی و همچنین خصوصی خود درباره ی موضوعات مختلف هستند. توییت یکی از این شبکه های اجتماعی است که در دهه اخیر محبوبیت بسیاری یافته است. این شبکه اجتماعی روشی سریع و موثر برای تحلیل احساسات، دیدگاهها و انتقادات مشتریان برای موفقیت در بازار را به سازمانها ارائه می دهد. تحلیل احساسات یا عقیده کاوی فرآیندی است که در آن نظرات، احساسات و نگرش افراد در ارتباط با موضوعی خاص استخراج می شود. پژوهش های زیادی در رابطه با تحلیل احساسات بر روی نظرات کاربران، مستندات و مقالات انجام شده است. تحلیل بر روی موارد بیان شده تفاوت عمده ای با داده های توییت دارد، به این سبب که توییت های توییت محدودیت 280 کاراکتری دارند و کاربران را وادار به بیان احساسات خود به صورت فشرده و کوتاه می نمایند. بهترین نتایج به دست آمده در طبقه بندی احساسات از تکنیک های یادگیری ماشین مثل بیز ساده و ماشین بردار پشتیبان حاصل شده است. در این پژوهش به ارائه روشی برای تحلیل احساسات در شبکه های اجتماعی پرداخته می شود. در این راستا سعی شده با تمرکز بر مراحل پیش پردازش داده ها و انتخاب ویژگی، طبقه بندی متن توسط روش بیز را تا حدودی بهبود بخشیم. به عبارتی، با تعریف این مسئله به صورت یک مسئله کلاس بندی باینری بر اساس خصیصه های پیشنهادی به تحلیل احساسات کاربران پرداخته می شود. مسئله کلاس بندی با استفاده از جدیدترین دستاوردهای حوزه یادگیری ماشین فرموله و حل شده است. برای ارزیابی روش پیشنهادی در این رساله از سناریو مجموعه دادگان توییت می باشد. روش پیشنهادی با سایر روش های طبقه بندی مقایسه می شود. بهترین عملکرد را از خود نشان داده است.

واژه های اصلی: شبکه های اجتماعی، تحلیل احساسات، محتوا، یادگیری ماشین، کاربران.

1- مقدمه

که شرکتها بهتر است خدمات خود را در قالب رسانه های اجتماعی ارائه نمایند و بیش از 85٪ معتقدند که کارخانجات باید تعامل بهتری از طریق رسانه های اجتماعی با مشتریان خود داشته باشند [13]. این آمار و ارقام نشان دهنده این موضوع است که چقدر رسانه های اجتماعی و ارائه سرویس بر روی این سامانه ها برای مشتریان اهمیت دارد و کسب و کارها برای افزایش سهم بازار و پیشرو شدن در رقبا بر روی این پروژه ها سرمایه گذاری می کنند. استفاده از رسانه های اجتماعی برای فرآیندهای مرتبط با مشتریان، که به آن مدیریت ارتباط با مشتری اجتماعی گفته می شود، بخش بزرگ تری از استفاده صرف کسب و کارها از رسانه های اجتماعی را شامل می شود. در اکثر منابع، مدیریت ارتباط با مشتری به چهار شکل کاملاً متفاوت مورد توجه قرار گرفته است: استراتژیک، عملیاتی، تحلیلی و تعاملی. در نگاه استراتژیک به مدیریت ارتباط با مشتری، CRM یک استراتژی محوری و کلیدی در کسب و کار است و

رسانه های اجتماعی با سرعتی تصاعدی در حال گسترش هستند و در حال حاضر آن ها فرصتی بی نظیر برای کسب و کارها ایجاد کرده اند. فیس بوک، به عنوان بزرگترین شبکه اجتماعی در حال حاضر بیش از 900 میلیون کاربر فعال دارد **Error! Reference source not found.** بیش از 70٪ این کاربر به صورت روزانه وارد این سایت شده و در حدود 940 میلیارد دقیقه را صرف بازدید از این سایت می نمایند. همین وضعیت برای دیگر رسانه های اجتماعی نظیر توییت صادق است. با شیوع بیماری کرونا و همگیر شدن آن، کاربران از صنایع مختلف انتظار دارند سرویس و محصولات خود را در بستر آنلاین به فروش برسانند و آن ها بتوانند در هر شرایطی به محصولات مورد نیاز خود دسترسی داشته باشند. **Error! Reference source not found.** بر پایه تحقیق کن 2020 بیش از 98٪ کاربران آمریکایی رسانه های اجتماعی عقیده دارند

قرار است مشتریان سود ده را جذب کسبوکار کرده و آن‌ها را برای ما حفظ کند. در نگاه عملیاتی

edu.myresearch@gmail.com

به مدیریت ارتباط با مشتری، هر فرایندی که به نوعی با مشتری مرتبط است. با استفاده از سامانه‌های نرم‌افزاری، به ابزارهای خودکار تجهیز می‌شود. بازاریابی، فروش و خدمات مشتریان، از جمله این فرایندها هستند. در نگاه تحلیلی به مدیریت ارتباط با مشتری، CRM ابزاری برای تحلیل هوشمندانه داده‌ها و اطلاعات مربوط به مشتری باهدف‌های استراتژیک و یا عملیاتی است. در نگاه تعاملی به مدیریت ارتباط با مشتری، CRM از فناوری برای مدیریت مرزهای سازمان، چه در رابطه با مشتری و چه شرکای تجاری استفاده می‌شود و هر نوع داده و اطلاعاتی که از مرزهای سازمان عبور می‌کند توسط این سیستم مدیریت می‌شود. هدف این سیستم ایجاد ارزش برای مشتریان و همکاران سازمان است. با هر دیدگاهی که به CRM توجه شود رسانه‌های اجتماعی می‌تواند جایگاه مهمی داشته باشد. یکی از زیر بخش‌های مهم برای تحقق SCRM، تحلیل احساسات و علایق مشتریان می‌باشد. با توجه به وسعت و قابلیت‌های تعامل کاربران، این بخش نیازمند تحول بنیادین با توجه به نیازهای روزافزون فضای کسبوکار و روش‌های بازاریابی می‌باشد [1].

تحلیل احساسات و علایق مشتریان قلب هر سامانه مدیریت ارتباط با مشتری محسوب می‌شود. مسئله مطرح در این پروژه، ارائه روشی برای تحلیل احساسات مشتریان در سامانه‌های مدیریت ارتباط با مشتری اجتماعی است. این روش قادر خواهد بود با بررسی پست‌ها و فعالیت‌های کاربران در رسانه‌های اجتماعی به شناسایی احساسات، اقدام به گروه‌بندی مشتریان برحسب علایق و احساسات نماید. همین‌طور این روش قادر است گزارش‌گیری کارآمدی از تغییرات نظرات و احساسات مشتریان در بازه‌های زمانی مختلف را ارائه نماید. تحلیل احساسات یکی از فعالترین حوزه‌های پژوهشی در پردازش زبان طبیعی است که به دلیل اهمیت آن در کسب و کار و جامعه به خارج از علم کامپیوتر مانند مدیریت و علوم اجتماعی نیز گسترش یافته است. اهمیت درحال رشد تحلیل احساسات با رشد رسانه‌های اجتماعی مانند بررسی ها، انجمن ها، بحث، وبلاگها، میکروبلگها، توییت‌ها و شبکه‌های اجتماعی همخوانی دارد. سیستم‌های سنجش در هر حوزه تجاری و اجتماعی کاربرد دارند؛ زیرا نظرات در همه احساسات تقریباً فعالیتهای انسانی متمرکز هستند و تأثیرات کلیدی رفتارهای ما به شمار می‌روند. کسب و کارها با تحلیل احساسات کاربران خود می‌توانند سرویس یا محصول خود را ارتقا دهند یا محصولات و سرویس‌های جدید را به بازار ارائه دهند. از طرفی تحلیل احساسات کاربران می‌تواند به تصمیمات مدیران سازمان جهت ارائه تبلیغات موثرتر، بهبود پرموشن‌ها و جذب کردن کاربران کمک کند. همچنین در صورتی که مشتریان نسبت به محصولات یا سرویس‌های فعلی احساس ناخوشایندی داشته باشند

می‌توان با دخیل کردن نظرات آن‌ها این حس ناخوشایند را به احساس مثبت و افزایش رضایتمندی آن‌ها تبدیل کرد.

در بخش دوم به ارائه کارهای مرتبط پرداخته می‌شود. در بخش سوم کار پیشنهادی مطرح می‌شود و در نهایت در بخش چهارم به ارزیابی روش پیشنهادی پرداخته می‌شود.

2- پیشینه تحقیق

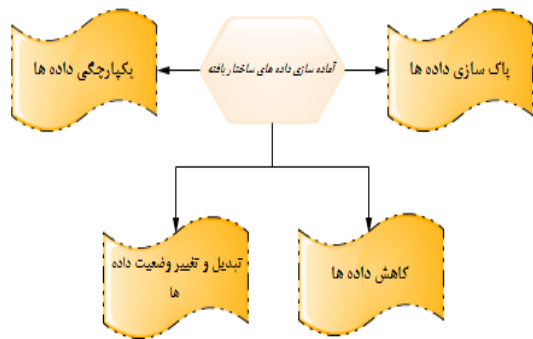
بیشتر مطالعات مرتبط با تحلیل احساسات در گذشته بر اساس الگوریتم‌های یادگیری بانظارت انجام گرفته است که نیاز به تهیه داده برچسب خورده دارند. مدل بیز ساده، ساده‌ترین و پراستفاده‌ترین الگوریتم احتمالاتی برای دسته بندی است و بر مبنای قضیه بیز کار میکند. این مدل احتمالات پسین رویدادها را محاسبه کرده و برچسبی که بیشترین احتمال پسین را دارد به رویداد نسبت میدهد. دسته بندی پرکاربرد آنروپی بیشینه است که کار دسته بندی را میتوان با آن انجام داد. این روش بر پایه مدل نمایی و اصل حداکثر آنروپی است. استفاده از این روش تجربه‌های موفق در کار پردازش زبان طبیعی از جمله در تحلیل احساسات به ارمغان آورده است. این روش در اکثر (و نه در همه) مواقع نسبت به مدل بیز ساده برتری دارد. ماشین بردار پشتیبان کار دسته بندی اسناد بر مبنای موضوعات مشابه بسیار مفید است. روش SVM یک مدل یادگیری بانظارت است که کار آن دسته بندی کردن اشیاء در کلاسهای مختلف با استفاده از ویژگیهای استخراج شده است. این دسته بندی با ایجاد ابرصفحه‌های میان نمونه‌های هر کلاس و حداکثر کردن فاصله نمونه‌ها از این صفحه صورت می‌گیرد. برتری این روش نسبت به دیگر روشهای مطرح یادگیری ماشین آن است که در مورد داده‌های ورودی پیش فرضی ندارد و به جای تکیه بر ارزشهای احتمالاتی، سعی دارد تا بهترین دسته بندی را با داده‌های موجود انجام دهد و نتایج به دست آمده از آن در تحلیل احساسات برتری محسوس به دیگر روشهای یادگیری ماشین در زبان انگلیسی دارد [8]. در سالهای اخیر روشهای یادگیری عمیق به خصوص شبکه‌های (RNN) در تحلیل احساسات برای زبان انگلیسی چینی و آلمانی در میان زبانهای مختلف، با استفاده از بردارهای مختلف نمایش کلمات کاربرد زیادی داشته است. آنها برای درک و کنترل ترکیب معنایی در کارهای پیچیده‌های مانند تحلیل احساسات مفید هستند. شبکه‌های RNN برای داده‌هایی با قابلیت تبدیل به مقادیر متوالی به کار می‌روند و با استفاده از ایده اشتراک گذاری پارامترها برای رسیدن به وزنهای مطلوب، توانایی پردازش توالیهایی با طولهای متفاوت را دارند. با وجود این که استفاده از آنها در تحلیل احساسات برای زبان انگلیسی با نتایجی بهتر از روشهای یادگیری بانظارت همراه بوده است. با رشد ساختار شبکه‌های RNN، ابعاد ماتریسها در مرحله بازپخش به صورت توانی رشد میکنند و در عمل استفاده از آنها غیر ممکن میشود [11]. شبکه‌های پیچشی که کولوبرت و دیگران در ابتدا برای کاربرد در بینایی رایانه‌های ارائه کرده‌اند، اخیراً در بسیاری از کارهای

است با استفاده از تکنیک های زبان طبیعی انجام توصیف می کنند. [14] تقاضا برای اطلاعات فرابری شده از منابع متنی به طور فزاینده ای در حال افزایش است. ذات غیرساخت یافته ی این متون، اعمال همان روشهایی را که ما در مورد دیتابیس ها بکار می بریم، غیر ممکن می سازد. کاربردهای مهمی را که از پردازش متون مورد انتظار است، بررسی می کنیم. به اینگونه پردازشها که روی متون اعمال می شود، متن-کاوی می گوئیم [5]. وظیفه‌ی اصلی عقیده کاوی طبقه‌بندی قطبیت است. طبقه‌بندی قطبیت وقتی اتفاق می‌افتد که یک نکه متن که یک عقیده در مورد یک موضوع را بیان کند به یکی از دو احساس متضاد تقسیم شود. نظرانی مثل «موافق» در مقابل «مخالف»، «دوست داشتن» در مقابل «دوست نداشتن» مثال‌هایی از طبقه‌بندی عقاید هستند. طبقه‌بندی قطبیت بیانات موافق و مخالف را تشخیص می‌دهد و به تولید ارزیابی‌های معتدتر کمک می‌کند [2]. بسیاری از شرکت‌ها از عقیده کاوی و تحلیل احساسات به عنوان جزئی از تحقیقاتشان استفاده می‌کنند. مثلاً شرکت‌ها از عقیده کاوی برای ساخت و نگهداری نظرات استفاده می‌کنند. سیستم‌های آن‌ها به طور مداوم اطلاعات را از وب مثل نظرات محصولات، دریافت برند و مسائل سیاسی جمع آوری می‌کند. دیگر سیستم‌ها نیز ممکن است از عقیده کاوی و تحلیل احساسات به عنوان یک فناوری زیر مؤلفه برای بهبود مدیریت روابط مشتری و سیستم توصیه‌گر از طریق بازخوردهای مثبت و مشتریان استفاده کنند. به طور مشابه، عقیده کاوی و تحلیل احساسات ممکن است شعله‌ها (زبان خصومت‌آمیز و گرمای اضافی) را در روابط اجتماعی شناسایی و حذف کنند [12]. طبقه‌بندی بیان احساسات بر اساس معنای آن‌ها و دانش قبلی تمایلات معنایی نامیده می‌شود. با وجود اینکه تحلیل نحوی نقشی کلیدی در طبقه‌بندی اسناد بازی می‌کند اما این برای استخراج مفاهیم از متن فقط از طریق نحو کافی نیست. معیارهای تئوری اطلاعات و دانش معنایی یک سلسله مراتب را با استفاده از WordNet ترکیب کردند تا مفاهیم را به طور اتوماتیک از متن استخراج کنند [15] دنکه نقش مدل بر پایه قوانین و یادگیری ماشین در یک دامنه چندگانه، بر روی سناریو طبقه بندی تست کرده است نتایج آنها نشان می دهد که رویکرد مبتنی بر واژگان، که با استفاده از SentiWordNet ساخته شده است، دقت آن در مقایسه با روش های یادگیری ماشین محدود تر است [10].

3- روش پیشنهادی

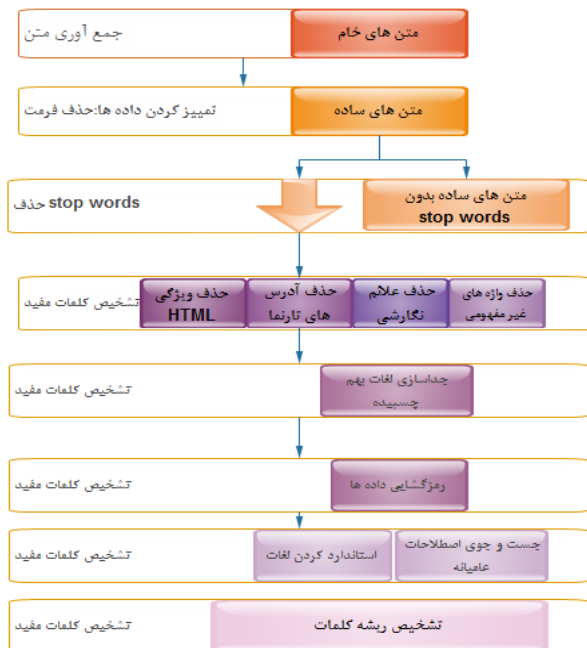
در این قسمت به ارائه روش پیشنهادی برای تحلیل احساسات مشتریان در رسانه اجتماعی خواهیم پرداخت. این روش شامل پنج مرحله به صورت شکل می باشد. این مراحل شامل جمع آوری داده، پیش پردازش داده، آماده سازی داده، برچسب گذاری کلمات، شناسایی خصیصه های مرتبط با دامنه و خوشه بندی کلمات مرتبط با احساسات می باشد. این مراحل به صورت تفصیلی در ادامه بخش تشریح خواهد شد.

پردازش زبان طبیعی مانند تجزیه نحوی، تجزیه سطحی، برچسب زنی نقش معنایی مورد استفاده قرار گرفته است. استفاده از شبکه های پیچشی قطعه بندی در تحلیل احساسات نیز برای زبانها با منابع فراوان مورد استفاده قرار گرفته و باعث بهبود قابل توجه دقت و کاهش زمان مرحله آموزش نسبت به دیگر روشهای یادگیری عمیق شده است. پژوهشهای حوزه تحلیل احساسات در زبان فارسی معمولاً یا با استفاده از روشهای مبتنی بر قاعده هستند یا مبتنی بر پیکره. برای بهبود نتایج معمولاً از پیش‌پردازش نظرات و ویژگیهای لغتنامه استفاده شده است. بصیری و همکاران [6] یک چارچوب مبتنی بر لغتنامه ارائه کردند که به صورت بدون نظارت با استفاده از قواعد از پیش تعیین شده و لغتنامه تعریف شده جهتگیری متون محاوره را تشخیص میدهد. استفاده از SVM برای تحلیل احساسات در زبان فارسی بر روی داده مربوط به نقد فیلم، منجر به نتایج بهتری نسبت به روشهای دیگر یادگیری ماشین شده است. بازدهی این روشها وابسته به کیفیت برچسبدهی درپیکرها و شیوه گزینش ویژگیها پیش از شروع کار دسته بندی است. روشنفکر و همکاران [7] برای اولین بار از شبکه های عصبی LSTM برای تشخیص احساسات متون فارسی استفاده کردند و توانستند نسبت به روشهای یادگیری سنتی نتایج بهتری داشته باشند، اما این نوع شبکه ها برای آموزش نیاز به دادههای خیلی زیادی دارند. همچنین آنها در کار خود فقط دو سطح از احساس را در نظر گرفتند و از جاسازی ساده کلمات استفاده کردند. اینترنت از قابلیت‌ها و امکانات زیادی برای ایفای کارکرد حوزه عمومی برخوردار است. شبکه اینترنت امکاناتی در اختیار مردم جوامع می‌گذارد تا در فضایی مناسب به گفتگوی آزاد و برابر با هم بپردازند و در نتیجه فرآیندهای گفتگو و مباحثه، به نقطه‌نظرهای مشترکی درباره مسائل سیاسی و اجتماعی دست یابند و به افکار عمومی شکل دهند. (میناوند، محمدقلی) مجموعه داده های نشات گرفته از تحقیق بر روی شبکه های اجتماعی در زمینه های بسیاری مانند جامعه شناسی و روانشناسی با ارزش هستند. اما حمایت از دیدگاه فنی به اندازه کافی دور است، و به روشهای خاص فوری نیاز دارند. (اکسیو وانگ و همکار) تحقیق روانشناسی برای تشخیص کاربران افسرده شبکه های اجتماعی [16] را با استفاده از داده کاوی انجام داده است. اخیراً ایده استفاده از تجزیه و تحلیل احساسات کاربران شبکه های اجتماعی برای بهبود عملکرد برنامه های کاربردی در وب سایت های خرید آنلاین توجه پژوهشگران را به خود جلب کرده است. وب سایت های خرید آنلاین بطور گسترده ای بررسی در مورد یک محصول ارائه می کنند و مشتریان میتوانند استفاده کنند. [4] در اینجا تجزیه و تحلیل احساسات مشتریان در باره هر محصول انجام می شود. در بررسی از سایت های شبکه اجتماعی قوانین برای احساسات مثبت یا منفی بسته به نمره کلی آن، با کمک SentiWordNet محاسبه می شود. تجزیه و تحلیل تمایلات شامل تشخیص ذهنیت و احساسات موجود در نظرات است. نظرات عبارات توصیف عواطف و احساسات مردم در مورد یک موضوع، نهاد و یا رویداد

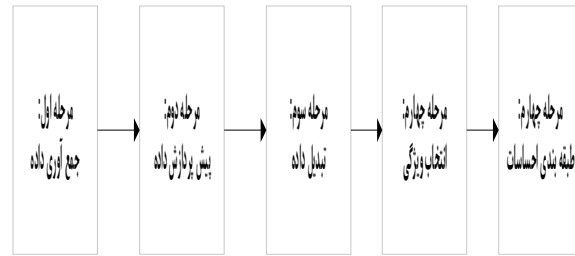


شکل (2): آماده سازی داده های ساختار یافته

آماده سازی داده های نیمه ساختاریافته و غیر ساختار یافته - داده های نیمه ساختاریافته شکلی از داده های ساختاریافته ای هستند که از ساختار رسمی از جداول و مدل های داده ای وابسته به پایگاه داده های رابطه ای مطابقت نمی کنند، اما با این وجود شامل برچسبها یا علامتها و شاخص هایی هستند که عناصر معنایی را از یکدیگر جدا می کنند و سلسله مراتبی از رکوردها و فیلدها را بین داده ها ایجاد می کنند. داده های غیر ساختار یافته به داده هایی گفته می شود که از هیچ مدل از قبل تعریف شده ای تبعیت نمی کنند مثالی از داده های ساختاریافته متن های سنگین می باشد.



شکل (3): آماده سازی داده های نیمه ساختار یافته و غیر ساختار یافته



شکل (1): روش پیشنهادی

مرحله اول، ورودی نظرات کاربران

کاربران با ثبت نام بر روی شبکه های اجتماعی، فعالیت خود را بر بستر این شبکه ها آغاز می کنند. این فعالیتها شامل تولید محتوا توسط کاربران، پیگیری علاقمندی هایشان، عضو شدن در گروه های متنوع و غیره می باشد. تمامی فعالیت های انجام شده توسط این کاربران در پایگاه داده ذخیره می شود، بنابراین انواع مختلفی از داده های متنی یا غیر متنی، داده هایی با ساختارهای متنوع، داده های غیر دقیق و نادرست وجود دارد. در مراحل بعدی سعی در تمییز کردن داده ها می شود. در مرحله نخست به جمع آوری داده پرداخته می شود. داده هایی می توانند از رسانه های مختلف جمع آوری شود. در این پایان نامه داده توئیتر انتخاب شده است. برای جمع آوری داده های توئیتر چندین ابزار کاربردی متفاوت، وجود دارد. (1) برنامه کاربردی جستجوی توئیتر، (2) برنامه کاربردی جریان توئیتر، (3) ابزار آتش نشانی توئیتر.

مرحله دوم، پیش پردازش داده ها: جداسازی جملات

مشتریان با ثبت نام بر روی رسانه های اجتماعی، فعالیت های خود را بر بستر این شبکه ها آغاز می کنند. این فعالیتها شامل تولید محتوا توسط کاربران، پیگیری علاقمندی هایشان، عضو شدن در گروه های متنوع، ارائه نظرات مثبت و منفی خود نسبت به یک محصول یا سرویس خاص می باشد. تمامی فعالیت های انجام شده توسط این مشتریان در پایگاه داده ذخیره می شود، بنابراین انواع مختلفی از داده های متنی یا غیر متنی، داده هایی با ساختارهای متنوع، داده های غیر دقیق و نادرست وجود دارد. در این مرحله برای از بین بردن عدم این ناسازگاری ها، کمبودهای یکپارچگی و بهبود تحلیل نتایج به آماده سازی داده ها پرداخته می شود، برای این کار داده ها، به سه دسته ساختار یافته، نیمه ساختار یافته و غیر ساختار یافته تقسیم می شود.

آماده سازی داده های ساختار یافته

داده های ساختار یافته از درجه بالای سازمان دهی برخوردار هستند. برای آماده سازی داده های ساختار یافته چهار گام اصلی که در شکل 2 مشاهده می شود، طی خواهد شد.

مرحله سوم: تبدیل داده

در این بخش داده ها را به فرمتی مناسب تبدیل کرده و آن ها برای مرحله بعدی آماده می شوند.

گاهی داده های خامی که برای تحلیل داریم مناسب گروهی از آزمون های آماری نیستند و برای این که بتوانیم از این دسته آزمون های آماری استفاده کنیم و همچنین دقت تحلیل را بالا ببریم باید در داده های خام تغییراتی ایجاد کنیم. یکی از این تغییرات، تبدیل داده ها نام دارد. تبدیل داده ها، روش هایی ریاضی است که برای تعدیل متغیرهایی به کار می رود که از مفروضه های آماری نرمال بودن، خطی بودن و یکسانی پراکندگی پیروی نمی کنند یا الگوهایی با داده های پرت غیرمعمول دارند

در مجموع زمانی که پیش شرط های آزمون های چندمتغیره برقرار نباشد، باید داده های به دست آمده را تبدیل کنیم تا امکان استفاده از برخی آزمون های مدنظر (عموما پارامتریک) فراهم شود.

در ابتدا باید میزان تخطی و تفاوت داده ها از پیش فرض های ذکر شده را تعیین کرد و در صورتی که پیش فرض ها یا پیش شرط های آماری به دست آمده دارای تفاوت قابل اعتنایی با مقدار معیار باشند از روش تبدیل داده ها استفاده کرد. تبدیل داده ها با هدف تعدیل متغیرها از جنبه علمی روشی پذیرفته شده است. البته زمانی که اختلاف داده ها با پیش فرض های آماری اندک باشد و به طور تقریبی مفروضات آماری برقرار باشد می توان از تبدیل داده ها صرف نظر کرد.

باید توجه داشت که تبدیل داده ها تا اندازه ای مانند شمشیر دولبه است. حسن این روش این است که می تواند دقت معنی داری تحلیل های آماری را افزایش دهد و عیب آن این است که ممکن است تفسیر داده ها را دشوارتر کند. در نتیجه باید از روش تبدیل داده ها به شیوه ای مدبرانه استفاده کرد.

دشواری در تفسیر داده ها بدین معناست که وقتی داده ها را تبدیل می کنیم، مقدار حداقل و حداکثر و شیوه توزیع متغیر و تمامی شاخص های میانگین و انحراف استاندارد تغییر می کند و با حالت معمول و عادی تفاوت پیدا می کند. مثلا اگر سن افراد که به صورت کمی (نسبی) سنجیده شده است را به توان دو برسانیم شاخص های آماری سن افراد تغییر می کند و با سن های غیر عادی مثل 250، 300 و غیره مواجه می شویم. یا وقتی متغیری مانند اعتماد اجتماعی داریم و با 10 سوال این متغیر را سنجیدیم و دامنه میانگین این متغیر بین 1 تا 5 باشد، لگاریتم گرفتن از این متغیر دامنه نمرات را تغییر می دهد و توضیح و تفسیر متغیر را با مشکل مواجه می کند. یکی از راه های رفع این مشکل این است که هنگام گزارش یافته های توصیفی و شاخص های آماری (مانند میانگین، انحراف استاندارد و مقدار حداقل و حداکثر) یافته ها و شاخص های آماری را هم به صورت عادی (قبل از تبدیل داده ها) و هم بعد از تبدیل داده ها گزارش کنیم.

مرحله چهارم: انتخاب ویژگی ها

در این مرحله به کلمات وزن داده می شود. کلمات با توجه به وزنی که در این مرحله و با استفاده از معیارهای تعریف شده در مرحله بعدی دریافت می کنند، مشخص می شوند. نام دیگر این مرحله استخراج ویژگی است. یعنی در این مرحله ویژگی های موردنظر را مشخص می کنیم تا در مرحله بعدی انتخاب شوند

در مرحله انتخاب ویژگی، ویژگی هایی که معیارهای تعریف شده را ماکزیم می کنند، انتخاب خواهند شد. این روش با کاهش نمونه ها سعی در ایجاد یک دسته بندی مناسب دارد.

-تشخیص توییت های مورد علاقه کاربر

اولا بایستی مرتبط بودن توییت را با علایق کاربر پیدا کرد. به منظور تعیین میزان ارتباط یک توییت با موضوعات مورد علاقه کاربر در اغلب کارهای انجام شده از روش TF-IDF و کسینوس زاویه بین بردار کلمات tw_i و pu_i استفاده می گردد. DF این معیار تعداد تکرار ویژگی ها را در متون بررسی می کند. در این معیار یک میزان آستانه تعریف می شود. آن ویژگی هایی که تعداد تکرارشان بیشتر یا کمتر از میزان آستانه است، حذف می شوند. تکرار زیاد یک ویژگی در اینجا احتمال انتخاب آن ویژگی را بیشتر می کند. از جمله ویژگی های این روش، مقیاس پذیری، سادگی و تأثیر آن است. معیار دیگر مورد استفاده "بهره اطلاعاتی" می باشد. این روش با استفاده از آنتروپی قابل محاسبه است. و آن ویژگی هایی که میزان gain بیشتری دارند را انتخاب می کند. برای افزایش دقت روش پیشنهادی از آنتولوژی استفاده می گردد. آنتولوژی برای مدل سازی شرایط در یک دامنه مورد علاقه و همچنین روابط میان این شرایط استفاده می شود. مهم ترین بخش آنتولوژی نقش کلیدی آن در توسعه وب معنایی است. تحلیل احساسات با استفاده از آنتولوژی به این صورت است که از آنتولوژی جهت استخراج مفاهیم مرتبط استفاده می شود. این بخش به بخش استخراج ویژگی اعمال شده است (در شرح مسئله، بخش استخراج ویژگی ها توضیح داده شده است). در واقع آنتولوژی یک نوع معنانشناسی انجام می دهد. کلمات یکسان، ممکن است معانی مختلف و کلمات مختلف، ممکن است معانی یکسان داشته باشند. آنتولوژی کلماتی را که از نظر مفهوم به آن ویژگی ها نزدیک ترند، مشخص می کند.

-بررسی تعداد لایک، تعداد کامنت ها و تعداد ذکر شدن ها:

لایک شدن یک پست توسط کاربر نشان دهنده آن است که کاربر نسبت به آن موضوع حساسیت بیشتری دارد. بنابراین تعداد لایک می تواند پارامتر مهمی در تشخیص احساس کاربر باشد. از طرفی اگر تعداد منشن ها زیاد باشد یعنی کاربران تمایل دارند موضوع را با سایر دوستان خود به اشتراک بگذارند. و همینطور تعداد کامنت ها نشان دهنده آن است که موضوع برای کاربر جذاب بوده و به ارائه ایده خود پرداخته است.

مرحله پنجم: کلاس بندی

جمع آوری شده است. برای این مقاله از سه موضوع مختلف ، کوید 19 ، آتشسوزی در جنگل های استرالیا و بهار عرب استفاده شده است.

جدول (1) : جزئیات مجموعه داده

| شماره | مجموعه دادگان | تعداد پست ها | تعداد کاربران | تعداد بازنشرها |
|------------------|------------------------------|--------------|---------------|----------------|
| مجموعه دادگان یک | کوید 19 | 23457 | 65890 | 654780 |
| مجموعه دادگان دو | آتشسوزی در جنگل های استرالیا | 4500 | 78904 | 567890 |
| مجموعه دادگان سه | بهار عرب | 5498 | 456789 | 47890 |

دانشی که در مرحله یادگیری مدل تولید می شود، می بایست در مرحله ارزیابی مورد تحلیل قرار گیرد تا بتوان ارزش آن را تعیین نمود و در پی آن کارائی الگوریتم یادگیرنده مدل را نیز مشخص کرد. این معیارها را می توان هم برای مجموعه داده های آموزشی در مرحله یادگیری و هم برای مجموعه رکوردهای آزمایشی در مرحله ارزیابی محاسبه نمود.

جدول (2) : معیار ارزیابی

| | | نتایج پیش بینی | |
|-------------|---------------|----------------|----------------|
| | | درست positive | غلط negative |
| نتایج واقعی | درست positive | True Positive | True Negative |
| | غلط negative | False Positive | False Negative |

شبکه های عصبی پیچشی ردهای از شبکه های عصبی عمیق هستند که معمولاً برای انجام تحلیل های تصویری یا گفتاری در یادگیری ماشین استفاده می شوند. این شبکه یک الگوریتم یادگیری عمیق است که تصویر ورودی را دریافت می کند و به هر یک از اشیا/جمله های موجود در تصویر میزان اهمیت (وزن های قابل یادگیری و بایاس) تخصیص می دهد و قادر به متمایزسازی آن ها از یکدیگر است. در الگوریتم ConvNet در مقایسه با دیگر الگوریتم های دسته بندی به پیش پردازش کمتری نیاز است. در حالیکه فیلترهای روش های اولیه به صورت دستی مهندسی شده اند، شبکه عصبی پیچشی، با آموزش دیدن به اندازه کافی، توانایی فراگیری این فیلترها/مشخصات را کسب می کند. معماری ConvNet مشابه با الگوی اتصال نورون ها در مغز انسان است و از سازمان دهی قشر بصری در مغز الهام گرفته شده است. هر نورون به محرک ها تنها در منطقه محدودی از میدان بصری که تحت عنوان میدان تاثیر شناخته شده است پاسخ می دهد. یک مجموعه از این میدان ها برای پوشش دادن کل ناحیه بصری با یکدیگر هم پوشانی دارند. ConvNet قادر است به طور موفقیت آمیزی وابستگی های زمانی و فضایی را در یک تصویر با استفاده از فیلترهای مرتبط ثبت کند و همچنین، معماری فیلترگذاری بهتری را روی مجموعه داده تصویر به دلیل کاهش تعداد پارامترهای درگیر و استفاده مجدد از وزن ها انجام می دهد.

4-ارزیابی

در این بخش به ارزیابی روش پیشنهادی که برحسب مجموعه دادگان توپیر است می پردازیم.

4-1 مجموعه دادگان

برای ارزیابی روش پیشنهادی بر طبق مراحل زیر عمل کرده و خروجی روش پیشنهادی با سایر الگوریتم های یادگیری ماشین مقایسه خواهد شد. برای جمع آوری داده در توپیر از روش جریان توپیر استفاده شده است. بر طبق این ابزار توپیرت های مورد نظر بر حسب موضوعات مختلف جدول (3) : روابط روش های مختلف ارزیابی

| | |
|---|---------|
| $Precision = \frac{tp}{tp + fp}$ | رابطه 1 |
| $Recall = \frac{tp}{tp + fn}$ | رابطه 2 |
| $F = 2 \cdot \frac{precision \cdot recall}{precision + recall}$ | رابطه 3 |
| $Accuracy = \frac{tp + tn}{tp + tn + fp + fn}$ | رابطه 4 |

علی رغم کندی آن می تواند قابلیت های این روش را نشان دهد. دقت بالای روش SASM به ماهیت این روش برمی گردد. در میان روش های دیگر، روش SVM عملکرد مناسبی داشته است. علت این برتری در قدرت طبقه بندی این روش می باشد.

در جدول 4 مقادیر دقیق حساسیت روش های مختلف ارزیابی شده را نشان می دهد. در بررسی نتایج، نکته حائز اهمیت دقت بالای روش SASM پیشنهاد شده در مقایسه با دیگر روش هاست. در SASM در برابر روشی مانند KNN بسیار کند است ، با این وجود، دقت بالای این روش

جدول (4): مقایسه نرخ حساسیت روش‌های پیشنهادی با بقیه روش‌ها

| مجموعه دادگان سوم | مجموعه دادگان دوم | مجموعه دادگان اول | میانگین | معیار حساسیت |
|-------------------|-------------------|-------------------|--------------|--------------|
| 0/786530612 | 0/88125000 | 0/896666700 | 0/854815771 | SASM |
| 0/484057971 | 0/456785714 | 0/416363600 | 0/4524024280 | KNN |
| 0/539074074 | 0/544255319 | 0/516483516 | 0/533270970 | نیویزین |
| 0/7045057971 | 0/778571400 | 0/696363600 | 0/726330990 | SVM |

تشخیص نیز، روش SVM به صورت میانگین، بهترین عملکرد را بعد از SASM به خود اختصاص داده است. روش SASM توانسته است، موارد عدم بازنشر را بهتر از روش‌های دیگر، تشخیص دهد. در این بخش، روش KNN بدترین عملکرد را به خود اختصاص داده است.

در جدول 5، روش‌های پیشنهادی با سه روش دیگر از نظر معیار تشخیص مقایسه شده است. در این ارزیابی، میزان پیش‌بینی درست تصمیمات عدم بازنشر، در معیار تشخیص بسیار موثر است. همان‌طور که در جدول 5، مشاهده می‌شود، SASM بهترین عملکرد را داشته است. در معیار

جدول (5): مقایسه معیار تشخیص روش‌های پیشنهادی با بقیه روش‌ها

| مجموعه دادگان سوم | مجموعه دادگان دوم | مجموعه دادگان اول | میانگین | معیار تشخیص |
|-------------------|-------------------|-------------------|-------------|-------------|
| 0/76407767 | 0/78969700 | 0/66073170 | 0/738168790 | SASM |
| 0/42777778 | 0/39532710 | 0/35934070 | 0/394148527 | KNN |
| 0/57317073 | 0/565591398 | 0/51031250 | 0/549691543 | نیویزین |
| 0/73015873 | 0/58571430 | 0/45918370 | 0/591668557 | SVM |

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 6 آورده شده است.

در ادامه ارزیابی روش‌های پیشنهادی برای پیش‌بینی تصمیم بازنشر کاربر در مواجهه با یک پست، به بررسی معیار نرخ خطای روش‌های پیشنهادی در مقایسه با روش‌های دیگر می‌پردازیم. معیار دقت، یکی از پرکاربردترین معیارها برای ارزیابی روش‌های پیش‌بینی و کلاس‌بندی است. هر چه

جدول (6): مقایسه نرخ دقت روش‌های پیشنهادی با بقیه روش‌ها

| مجموعه دادگان سوم | مجموعه دادگان دوم | مجموعه دادگان اول | میانگین | معیار دقت |
|-------------------|-------------------|-------------------|-------------|-----------|
| 0/752631579 | 0/8615385 | 0/68585710 | 0/766675726 | SASM |
| 0/389705882 | 0/48571430 | 0/34367820 | 0/406366127 | KNN |
| 0/562272730 | 0/691752577 | 0/682121212 | 0/645382173 | نیویزین |
| 0/739220779 | 0/78873240 | 0/59523810 | 0/707730426 | SVM |

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 سلیقه شوند. این

همان‌طور که در جدول 6 مشاهده می‌شود، روش SASM در این معیار، بهترین عملکرد را داشته است. بعد از این روش ماشین بردار پشتیبان

$(1=\beta)$ همان طور که اشاره شد، میزان این معیار، ارتباط مستقیمی به دقت و بازیابی روش ها دارد. هر دو روش پیشنهادی، بهترین عملکرد را داشته اند. نکته جالب در این جدول، پایین بودن مقدار F-measure برای روش KNN است. با وجود دقت نسبتاً خوب این روش، با توجه به ضعف شدید این روش در بازیابی، بدترین عملکرد را در این معیار دارد.

برتری روش پیشنهادی در دقت پیش بینی های انجام شده بسیار موثر است. در جدول 7، معیار F-measure روش های مختلف مقایسه شده است. همان طور که پیش تر اشاره شد، معیار F-measure بکارگرفته شده در این رساله، میانگین موزون دقت و بازیابی (حساسیت) می باشد

جدول (7): مقایسه معیار F measure روش های پیشنهادی با بقیه روش ها

| مجموعه دادگان سوم | مجموعه دادگان دوم | مجموعه دادگان اول | میانگین | معیار F measure |
|-------------------|-------------------|-------------------|-------------|-----------------|
| 2/579591837 | 2/283775 | 2/3 | 2/387780612 | SASM |
| 1/655347391 | 1/1857143 | 1/6990909 | 1/513384197 | KNN |
| 1/556701030 | 1/441124127 | 1/597011299 | 1/531612152 | نیویزین |
| 2/195121951 | 1/4482759 | 1/7142857 | 1/785894517 | SVM |

صورت پذیرد. به عنوان پیشنهاد برای کارهای آینده، می توان روی رفتارشناسی انتشار در شبکه های اجتماعی مطالعات تکمیلی صورت پذیرد. کاربران تاثیرگذار بر روی سایر کاربران اکتشاف شوند، این کاربران به گونه ای هستند که بیشترین احساس مثبت در جذب کاربران برای خرید یک محصول به سایر کاربران می دهند و همینطور می توانند بیشترین احساس منفی را برای دفع کاربران برای خرید یک محصول را در شبکه های اجتماعی داشته باشند. پست های تاثیر گذاری که بیشترین احساس مثبت و منفی را برای جذب یا دفع به خرید یک محصول می انجامد را اکتشاف کرد. با اتخاذ روش های دیگر برای مدل پیشنهادی می توان آن را مقیاس پذیرتر نمود. با دخیل کردن ویژگی هایی همچون فرهنگ، نژاد، قومیت به مدل پیشنهادی می توان نتایج گسترده تری را کسب نمود. با ترکیب مدل های طبقه بندی و قراردادن وزن های متفاوت به ویژگی های تاثیر گذار، عملکرد روش پیشنهادی را افزایش داد. از طرفی علایق کاربران با گذشت زمان تغییر می کند، این تغییرات یا بستگی به برهه های خاص زمانی دارد، مانند اوایل سال جدید و یا بر اثر تغییر طبع کاربر با گذر زمان ایجاد می شود، با در نظر گرفتن این پویایی در شبکه های اجتماعی می توان روش پیشنهادی را انعطاف پذیرتر نمود.

منابع و ماخذ

- [1] Arora, L., Singh, P., Bhatt, V., Sharma, B. (2021). **Understanding and Managing Customer Engagement through Social Customer Relationship Management**. Journal of Decision Systems, 1-21
- [2] Bagheri A., Sarraee, M. (2014). **Persian Sentiment Analyzer: a Framework Based on a Novel Feature Selection Method**. International Journal of Artificial Intelligence™, Vol. 12, No. 2, pp. 115-129.
- [3] Chen, Z. S., Zhang, X., Govindan, K., Wang, X. J., Chin, K. S. (2021). **Third-Party Reverses Logistics Provider Selection: A Computational Semantic Analysis-Based Multi-Perspective Multi-Attribute Decision-Making Approach**. Expert Systems with Applications, 166, 114051.

5- نتیجه گیری

تحلیل انتشار اطلاعات و نفوذ اجتماعی در شبکه های اجتماعی دارای کاربردهای بسیار زیادی در جهان واقعی دارد. یکی از مثال های کاربردی آن بیسنه سازی نفوذ در بازاریابی و پروسی می باشد. تعیین کاربران تأثیر گذار به عنوان یکی از اصلی ترین موضوعات موجود در شبکه های اجتماعی می باشد که اهمیت فراوانی دارد. چنانچه این کاربران به صورت دقیق تری شناسایی شوند، عملیاتی که بر مبنای این کاربران انجام می شود با نفوذتر خواهد بود. هدف از انجام این پروژه ارائه روشی برای تحلیل احساسات مشتری در رسانه های اجتماعی جهت استفاده در سامانه های تبلیغات است. برای این امر از API جست و جوی تویتر استفاده شده است. این مجموعه داده شامل اطلاعات مرتبط با کاربران و فعالیت هایشان در بستر شبکه اجتماعی تویتر بوده است که در پایگاه داده ذخیره شده است. این داده های ذخیره شده برای آماده سازی به سه دسته ساختار یافته، نیمه ساختار یافته و غیر ساختار یافته تقسیم شده است و برای هر کدام از ساختارها مراحل برای آماده سازی داده انجام شده است. سپس داده ها به فرمت مناسب تبدیل شده است و ویژگی های مورد نظر استخراج می شود و در نهایت به طبقه بندی کاربران بر اساس احساسات مثبت و منفی با استفاده از روش شبکه عصبی چرخشی خواهیم پرداخت. این روش با سه روش دیگر نیز برای ارزیابی مقایسه شد و از سه روش دیگر عملکرد بهتری داشت.

رفتارشناسی کاربران در شبکه های اجتماعی یکی از جذاب ترین بحث های حوزه فناوری اطلاعات در دهه اخیر می باشد. رفتارشناسی کاربران این امکان را در اختیار توسعه دهندگان فناوری اطلاعات فراهم می آورد که با استفاده از نیازسنجی تعاملات و برهم کنش علایق و خصوصیات کاربران، به ارائه سرویس بپردازند. این سرویس ها ممکن است در قالب ارائه آگهی تبلیغاتی، یا یک فرآیند مدیریت دانش و یا حتی به صورت یک توصیه

- [4] Cieliebak, M., Deriu, J., Egger, D., Uzdilli, F. (2017). **A Twitter Corpus and Benchmark Resources for German Sentiment Analysis**. in Proc of the 5th Ine, Workshop on Natural Language Processing for Social Media, SocialNLP, pp. 45-51, Boston, USA.
- [5] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P. (2011). **Natural Language Processing (Almost) from Scratch**. Journal of Machine Learning Research, vol. 12, No. 76, pp. 2493-2537.
- [6] Cortes C., Vapnik, V. (1995). **Support-Vector Networks**. Machine Learning, Vol. 20, No. 3, pp. 273-297.
- [7] Dos Santos C.N., Gatti, M. (2014). **Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts**. in Proc of the 25th International Conf. on Computational Linguistics, COLING'14, pp. 69-78, Dublin, Ireland, 25-29.
- [8] Jaynes, E.T. (1957). **Information Theory and Statistical Mechanics**. Physical Review, Vol. 106, No. 4, pp. 620.
- [9] Maulud, D. H. (2021). **State of Art for Semantic Analysis of Natural Language Processing**. Qubahan Academic Journal 1.2, 21-28.
- [10] Mikolov, T., I. Sutskever, I., Chen, K., Corrado, G.S., Dean, J. (2013). **Distributed Representations of Words and Phrases and their Compositionality**. in Proc Advances in Neural Information Processing Systems, NIPS'13, pp. 3111-3119, Lake Tahoe, CA, USA, 5-10.
- [11] Neethu M.S., Rajasree, R. (2013). **Sentiment Analysis in Twitter using Machine Learning Techniques**. in Proc IEEE 4th Int. Conf. on, Computing, Communications and Networking Technologies, ICCCNT'13, 5 pp., Tiruchengode, India, 4-6.
- [12] Roshanfekr, B., Khadivi, S., Rahmati, M. (2017). **Sentiment Analysis using Deep Learning on Persian Texts**. in Iranian Conf, on Electrical Engineering, ICEE'17, pp. 1503-1508, Tehran, Iran, 2-4.
- [13] Shearer, E., Amy, M. (2021). **News Use across Social Media Platforms in 2020**.
- [14] Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C. (2013). **Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank**. in Proc of the Conf.
- [15] Wang, K., Wang, X., Lin, L., Wang, M., Zuo, W. (2014). **3D human Activity Recognition with Reconfigurable Convolutional Neural Networks**. in Proc of the 22nd ACM International Conf. on Multimedia, pp. 97-106, Orland, FL, USA, 18-19.
- [16] Zhang, Y., Chen, M., Liu, L., Wang, Y. (2017). **An Effective Convolutional Neural Network Model for Chinese Sentiment Analysis**. in Proc AIP Conf. Proc., vol. 1836, pp. 020084, Rome, Italy, 27-29.
- [17] on Empirical Methods in Natural Language Processing, EMNLP'13, vol. 1631, pp. 1631-1642, Seattle, WA, USA, 18-21.