

ارائه مدلی برای تشخیص احساس از متن با رویکرد رای گیری اکثریت

علی جبار رشیدی^۱، کوروش داداش تبار احمدی^۲، ملک نظرانداز^۳

۱- دانشیار- دانشگاه صنعتی مالک اشتر- مجتمع دانشگاهی برق و کامپیوتر aiorashid@yahoo.com

۲- استادیار- دانشگاه صنعتی مالک اشتر- مجتمع دانشگاهی برق و کامپیوتر dadashtabar@yahoo.com

۳- دانشجوی کارشناسی ارشد رایانش امن - دانشگاه صنعتی مالک اشتر nazarandaz.m@gmail.com

چکیده

افراد در ارتباطات انسانی و درک نظرات معمولاً تکیه بر حقایق یا تعاریف شناخته شده دارند و پردازش‌ها را بر اساس آن انجام می‌دهند. نظرات حاوی اطلاعات ارزشمندی هستند که می‌توان با تحلیل احساسات آن‌ها به دانش ارزشمندی در ارتباط با یک موضوع خاص دست یافت چراکه تمرکز تحلیل احساسات بر روی تجزیه و تحلیل یک متن برای درک نظرات بیان شده است. در این تحقیق نیز از تحلیل احساسات مبتنی بر الگوریتم رای گیری اکثریت استفاده شد تا اطلاعات ارزشمندی از متن استخراج شود. هدف کلی مقاله بهبود روش‌های نظرکاوی با رای گیری اکثریت است. با کمک رویکرد ترکیبی واژگان و یادگیری ماشین یک سامانه تحلیل احساسات می‌تواند به طور موثر اطلاعات مرتبط با نظرات را به زبان طبیعی درک نماید. بسیاری از منابع از روش‌های مستقل برای نظرکاوی استفاده نموده‌اند که کمتر به ترکیب روش‌ها توجه می‌کند. با اجرای رویکرد ترکیبی و استفاده از مدل احساسی پلاتچیک بهبود یافته و انجام آزمایشات متعدد، کارایی و اثر بخشی مدل پیشنهادی در سطح جمله به دقت ۹۷٪ رسیده است.

کلمات کلیدی: تشخیص احساس، مدل احساسی پلاتچیک، شبکه‌های اجتماعی، رای گیری اکثریت

۱. مقدمه

رسانه‌های اجتماعی نحوه یادگیری افراد درباره وقایع فعلی را تغییر داده‌اند که منجر به تعامل جدید بین نویسندگان و خوانندگان شده است. رسانه‌های مرسوم مانند تلویزیون، رادیو و روزنامه‌ها حساب‌های فعالی در شبکه‌های اجتماعی دارند تا اخبار خود را به طور موثر برای جمعیت زیادی پخش کنند [۱]. توییتر یک مورد معروف در شبکه‌های اجتماعی برای انتشار اخبار است. این شبکه مورد علاقه رسانه‌ها برای ارسال اخبار شده است. به همین دلیل است که تقریباً ۸۵٪ از موضوعات پرتعداد عنوان خبری یا خبری ماندگار هستند [۲].

پست‌ها در شبکه‌های اجتماعی می‌تواند در مورد هر چیزی باشد. محدودیت‌های کمی در مورد موضوع یا نحوه بیان کاربران

وجود دارد. به همین دلیل، کاربران می‌توانند بدون هیچ‌گونه نگرانی صحبت کنند و نظرات خود را بیان کنند. این شکل خاص از ارتباطات باعث جلب توجه جامعه تجزیه و تحلیل احساسات شده است. زیرا هم نویسندگان و هم خوانندگان در طی مکالمه خود احساسات خود را ابراز می‌کنند. در واقع کاربرانی که روزگاری مصرف‌کننده اطلاعات بودند، اکنون همکاران اصلی اطلاعات هستند [۳]. علاوه بر این، محتوای پست‌ها معمولاً دارای محتوای احساسی بالایی است [۴] که در آن تشخیص دیدگاه‌های مثبت و منفی (قطبیت احساسات) و حتی تعیین احساس خاصی مانند شادی یا عصبانیت (واکنش احساسی) آسان است. در واقع تشخیص احساس فرآیند شناسایی احساس انسان است مثلاً برای موضوع تشخیص احساس اصولاً به شکل‌های انسانی (غیر ماشینی) نیمه خودکار (ماشین به کمک انسان) و خودکار (ماشین) قابل انجام است. در روش‌های نیمه خودکار و خودکار مسئله تعامل انسان-ماشین مطرح می‌شود که از پیچیدگی زیادی برخوردار است. که با ظهور شبکه‌های اجتماعی بخصوص در فضای مجازی با محوریت اینترنت مسئله‌ی استخراج احساس از صوت، متن، فیلم/تصویر و رسانه‌های ترکیبی محققان را مورد حوزه‌ی پژوهش پیچیده‌تری از گذشته نمود. طوری که بسیاری از ویژگی‌ها ساختار و ادبیات خاص شبکه‌های اجتماعی در بیان و تشخیص احساسات دارای نقش ویژه‌ای هستند بر این اساس تشخیص احساس در بستر شبکه‌های اجتماعی یکی از موضوعات تحقیقاتی روز دنیا می‌باشد. هدف از تحلیل احساسات، تعریف ابزارهای است که قادر به استخراج اطلاعات ذهنی از متون در زبان طبیعی، مانند نظرات و احساسات است تا توسط سیستم پشتیبانی یا تصمیم‌گیرنده برای ایجاد دانش سازمانی و عملی استفاده شود. در این مقاله روشی پیشنهاد شده است که با ارائه‌ی الگوریتم‌های مورد نظر با استفاده از روش رای‌گیری اکثریت [۵] [۶] به تحلیل احساسات در شبکه‌ی اجتماعی خواهیم پرداخت تا با استفاده از چالش‌های پیش‌رو به تحلیل و بررسی موضوع پردازیم. در بخش دوم این مقاله، به بررسی روش‌های مشابه پرداخته و در بخش سوم روش پیشنهادی استفاده شده توضیح داده شده است. در بخش چهارم، نتایج بیان گردیده و در نهایت در بخش آخر نتیجه‌گیری ارائه می‌شود.

۲. مروری بر ادبیات و کارهای مرتبط

جوشی و همکاران [۷] یک ردیاب احساسات منبع باز و برنامه‌های ابتکاری آن را ارائه می‌دهند. ردیاب، EmoGram، تغییرات احساسی را برای دنباله‌ای از واحدهای متنی ردیابی می‌کند. از نظر واحد متنی (توییت‌ها، جملات گفتگویی و غیره) و همچنین آنچه توالی زمان را تشکیل می‌دهد همه کاره است. با استفاده از واژگان EmoLex، LIWC، چهار احساس (خشم، مضطرب، خوشحال، غم) را مشخص می‌کند. بر طبق ادعای این پژوهش روش مورد استفاده می‌تواند با دقت ۶۷٪ احساس

کلی را در یک دوره زمانی پیش‌بینی کند. نمودارهای توالی زمانی احساسات را ایجاد و نشان می‌دهد. سطح احساس یک مسابقه کریکت، احساسات بین شخصیت‌های یک نمایش و احساسات نسبت به یک محصول را به طور دقیق محاسبه می‌کند. عادل مجید و همکاران [۸] بر روی تشخیص احساس از متن به زبان اردو رومی متمرکز است. یک مجموعه جامع جمله‌ای توسعه داده شده که از حوزه‌های مختلف جمع شده و آن را با شش کلاس (شاد، غم، خشم، ترس، عشق و خنثی) مختلف حاشیه نویسی می‌کنند. برای استخراج ویژگی از Word2Vec استفاده شده است. برای طبقه‌بندی از الگوریتم‌های مختلف پایه مانند کی-نزدیکترین همسایه، درخت تصمیم، ماشین بردار پشتیبان و جنگل تصادفی را بر روی مجموعه اعمال شد. پس از آزمایش و ارزیابی، بر طبق ادعای این پژوهش مدل ماشین بردار پشتیبان نسبت به بقیه‌ی الگوریتم‌های طبقه‌بندی با دقت ۶۹٫۵۴٪ به نتایج بهتری دست یافته است.

مارتینزو همکاران [۹] روش‌های طبقه‌بندی گفتار نفرت در رسانه‌های اجتماعی را بررسی می‌کنند. هدف این است که از روش

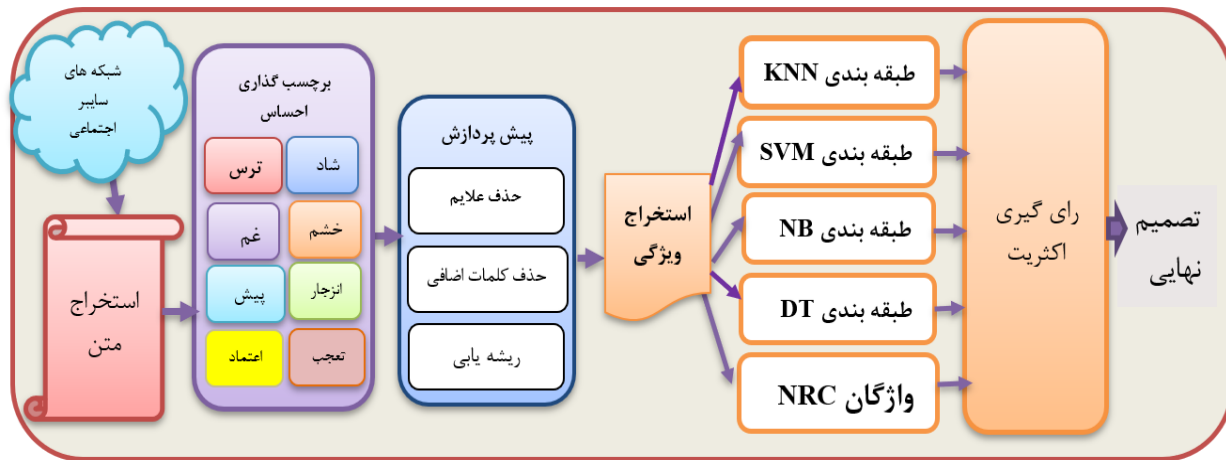
های طبقه بندی با استفاده از یک مجموعه داده حاشیه ای برای این منظور ، پایه های واژگانی را برای این کار ایجاد کنند. از TF-IDF برای مرتب کردن مناسب ترین کلمات در مجموعه داده استفاده شد به عنوان ویژگی ، سیستم از تکنیک های پردازش زبان طبیعی به منظور گسترش مجموعه داده اصلی با اطلاعات احساسی استفاده می کند از مدل پلاتچیک (ترس ، غم ، شادی ، خشم ، تعجب ، اعتماد ، پیش بینی ، انزجار) و از واژگان NRC به عنوان منبع ، تمام جملات مورد تجزیه و تحلیل قرار گرفته و شدت خشم احساس محاسبه می شود. که آن را برای طبقه بندی یادگیری ماشین فراهم می کند. با ادعای این پژوهش نتایج حاصل در شناسایی گفتار از متنفر با الگوریتم جنگل تصادفی (RF) بهترین دقت را دارد که ۵۶,۸۰٪ دقت بدست آوردند.

رشیدی و همکاران [۵] روش های ادغام تصمیم گیری برای داده های سنجش از دور را پیشنهاد نموده اند . در ادغام سطح تصمیم گیری ، نتایج دریافتی از طبقه بندی کننده محلی مختلف ترکیب می شود و تصمیم نهایی تعیین می شود. مهمترین روش های ادغام تصمیم گیری که در کاربردهای مختلف به کار رفته است و در این پژوهش نیز مورد استفاده قرار گرفته : تابع رای گیری اکثریت (MVF) ، بهترین روش رای گیری اکثریت (BMVF) ، روش مبتنی بر رتبه و روش مبتنی بر امتیاز، استنباط بیزی و روش دامپستر-شفراس. نتایج فیوژن با توجه به قابل قبول بودن و صحت طبقه بندی آنها با هم مقایسه می شود. مشکل اصلی و رایج در روشهای ادغام پیشنهادی ، ماهیت طبقه بندی محلی و شباهت کلاسهای داده است که نادیده گرفته شده اند و فقط به نتایج طبقه بندی محلی در یک نقطه (پیکسل) بسنده می شود ، بدون اینکه توجهی به توزیع نتیجه برای همه کلاس ها و سایر موارد پیکسل های مختلف داشته باشد. باتوجه به ادعای این پژوهش نتایج نشان می دهد که روش دمپستر-شفر دقیق تر از سایر روش های پیشنهادی است.

۳. روش پیشنهادی تشخیص احساسات اجتماعی در شبکه های سایبر اجتماعی

مسئله ی مورد استفاده در این مقاله ، طراحی یک مدل کارا برای کشف احساس اجتماع از متن است. با این حال ، برای مطالعه نظر یک کاربر ، جالب تر است که به طبقه بندی عمیق تر پردازیم با این وجود طبیعتاً احساس بسیار پیچیده است. یکی از رفتارهایی که نشان دهنده ی احساس یک فرد است، نحوه ی نوشتار او است ، که باتوجه به احساس بیان شده می توان احساسات پنهان شده در پشت پست وی را کشف کرد. همانطور که در جمع بندی پژوهش های پیشین بیان کردیم روش های ترکیبی از دقت و عملکرد بهتری نسبت به روش های مستقل عمل می کنند. بنابراین می توان گفت پژوهش حاضر بر طبق همین روش خواهد بود. در واقع روشی ارائه می شود که چگونه به یک ماشین آموزش دهیم تا تفاوت های تنوع فرهنگی را

تجزیه و تحلیل کند ، احساس را بیرون بکشد و احساس و معنی را در پشت کلمات با استفاده از تکنیک های یادگیری ماشین و واژگان پیدا کند. چارچوب کلی روش پیشنهادی بر طبق رای گیری اکثریت خواهد بود که با استفاده از آن به بهبود نتایج پرداخته خواهد شد.



شکل ۱- چهارچوب روش پیشنهادی

روش پیشنهادی ارائه شده شامل بخش‌های زیر می‌باشد که در ادامه هر کدام از این موارد توضیح داده خواهد شد.

۱.۳. شبکه‌های سایبر اجتماعی

شبکه‌های سایبر اجتماعی منبع غنی از اطلاعات در مورد عقاید هستند. شبکه‌های اجتماعی انواع مختلفی دارند که هر کدام نیز کارایی به خصوصی دارند. انواع شبکه‌های سایبر اجتماعی مثل فیسبوک، یوتیوب، اینستاگرام و توییتر و واتساپ و... می‌باشد اگر به ویژگی‌های یک پلتفرم اجتماعی نگاه کنیم، خواهیم دید که شبکه‌های اجتماعی به خوبی می‌توانند نقش یک پلتفرم موفق را ایفا کنند. یک پلتفرم موفق باید بتواند تمام نقاط جهان را در هر لحظه با کمترین سرعت و هزینه به هم متصل کند. حال آنکه شبکه‌های اجتماعی این کار را به خوبی انجام می‌دهند. پس می‌توان نتیجه گرفت شبکه‌های اجتماعی به عنوان یک پلتفرم اجتماعی صدای جهان را به همه می‌رسانند و افراد را به هم متصل می‌کنند.

۲.۳. استخراج متن

یکی از منابع اصلی برای جمع‌آوری اطلاعات در فضای سایبر شبکه‌های اجتماعی انواع پلت‌فرم‌ها است. شبکه‌های اجتماعی از لحاظ جمع‌آوری اطلاعات درجه اهمیت متفاوتی دارند. این اهمیت با توجه به نوع کشور نیز متغیر است چرا که در برخی از کشورها ممکن است اقبالی به یک شبکه‌ی اجتماعی خاص وجود نداشته باشد و بالعکس در کشور دیگری ممکن است مردم

علاقه‌مند به یک شبکه اجتماعی خاص باشند. استخراج متن بصورت دستی از شبکه‌های اجتماعی قابل انجام می‌باشد این کار می‌تواند با استفاده از خزنده‌ی وب (یک برنامه رایانه‌ای است که توانایی مرور و ثبت اطلاعات را از وبسایت‌ها به صورت خودکار دارد) نیز صورت پذیرد.

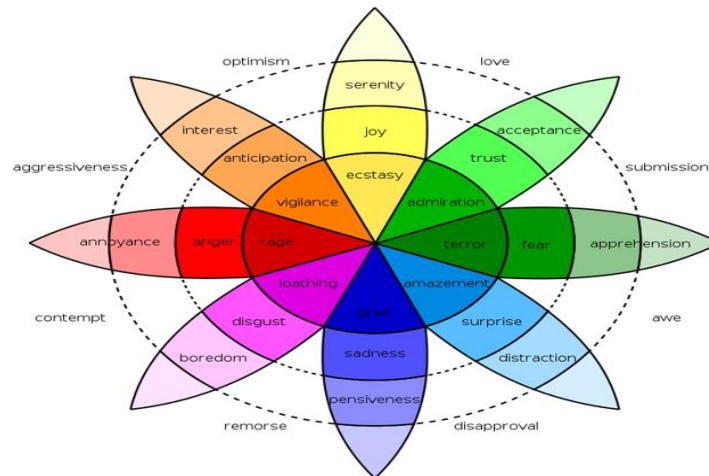


شکل ۲- نمونه ای از انواع شبکه های سایبر اجتماعی برای استخراج متن

۳.۳. بر چسب گذاری احساس

برچسب گذاری را نحوه آگاهی از مکالمات کاربران با تیم، تجزیه و تحلیل مکالمات برچسب گذاری شده و نهایتاً بهبود محصولات و خدمات می‌داند. برچسب گذاری ساختاریافته برای آموزش الگوریتم‌ها بسیار حائز اهمیت است. هنگامی که برچسب‌ها تعریف واضح و مشخصی نداشته باشند، انسان و ماشین هر دو سردرگم خواهند شد. سیستم های تشخیص احساس با توجه به نیازهای خود از مدل های مختلف احساس برای برچسب گذاری استفاده می کنند.

مدل احساسی که معرفی می شود، مدل پلاتچیک با ۳۲ احساس است. پذیرش، تحسین، پرخاشگری، حیرت، عصبانیت، دلخوری، انتظار، دلهره، هیبت، بی حوصلگی، تحقیر، عدم تأیید، انزجار، حواس پرتی، وجد، ترس، اندوه، علاقه، شادی، انزجار، عشق، خوش بینی، متفکر بودن، خشم، حسرت، غم، آرامش، تحریف، تعجب، وحشت، اعتماد و هوشیاری. هرچه دایره بزرگتر و به لبه چرخ نزدیکتر باشد، تجربه احساسی قوی تر خواهد بود. مدل پلاتچیک هشت احساس اصلی را بر اساس مبنای مثبت یا منفی گروه بندی می کند.



شکل ۳- مدل احساسی پلاتچیک

۴.۳. پیش پردازش

پیش پردازش داده‌ها وظیفه دارد اندازه مجموعه ویژگی‌ها را برای یادگیری مناسب کند. همان‌طور که از نام این روش پیداست، عملیات پیش‌پردازش معمولاً قبل از عملیات اصلی الگوریتم‌ها انجام می‌گیرند و باعث تسهیل و کمک به الگوریتم‌ها می‌شوند.

توکن‌یابی: به فرآیندی که طی آن جمله به یکسری واژه (توکن) تبدیل می‌شود. به عنوان مثال جمله‌ی زیر را در نظر بگیرید.
"تکلیف مشخصه و نیازی به توضیح نیست!"

با استفاده از توکن‌یابی میتوان به خروجی زیر دست یافت
'تکلیف'، 'مشخصه'، 'و'، 'نیازی'، 'به'، 'توضیح'، 'نیست'، '!'

حذف کلمات اضافی: در پردازش زبان طبیعی، کلمات بدون استفاده (غیر مفید در حین پردازش) را کلمات توقف می‌گویند. در واقع کلماتی هستند که به طور متداول استفاده می‌شوند. از جمله این کلمات متداول می‌توان به «از، به، و، را» و چنین مواردی در زبان فارسی اشاره کرد.

ریشه‌یابی: فرآیندی که طی آن ریشه یک واژه به دست می‌آید. مثل «کتاب‌ها» که دارای ریشه «کتاب» است یا «می‌خوردن» که دارای ریشه «خورد#خور» است.

۵.۳. استخراج ویژگی

پس از اتمام مرحله پیش‌پردازش، استخراج ویژگی برای ارزیابی داده‌های پردازش شده اعمال می‌شود. انتخاب و استخراج ویژگی مهمترین مرحله برای تشخیص احساسات است زیرا بر نتیجه کلی کار تأثیر می‌گذارد. یک انتخاب ویژگی خوب منجر به پیش‌بینی خوبی می‌شود. بنابراین، انتخاب صحیح ویژگی‌ها برای ارتقا طبقه‌بندی بسیار مهم است.

Unigram: (واژگان منفرد) در این مرحله ابتدا پاراگراف به جملات تقسیم می‌شوند و سپس جملات به صورت کلمه تبدیل می‌شوند.

Bigram: در این مرحله ابتدا پاراگراف به جملات تقسیم می‌شوند و سپس جملات به صورت واژگان دوتایی تبدیل می‌شوند.
TFIDF: هدف نشان دادن اهمیت کلمه کلیدی مورد نظر از طریق مقایسه تعداد تکرار کلمه در متن با تکرار آن کلمه در مجموعه‌ای بزرگ‌تر از مستندات می‌باشد.

Text To Sequences: این روش یک روش مبتنی بر جمله است که برای هر کلمه یک عدد در نظر گرفته می‌شود و سپس طول جمله‌های کوتاه با صفر پر می‌شود تا طول جمله‌ها با یکدیگر مساوی شوند تا بدین ترتیب یک ماتریس به اندازه n سند و m ویژگی بدست آید.

۶.۳. روش

الگوریتم مورد استفاده از نوع الگوریتم ماشین بردار پشتیبان (SVM)، بی‌س ساده (NB) و کی نزدیکترین همسایه (KNN) و درخت

تصمیم (DT) که از الگوریتم های نظارت شده در یادگیری ماشین هستند و همچنین از واژگان NRC نیز برای طبقه بندی احساس استفاده شده که در نهایت با استفاده از رای گیری اکثریت نتایج بهبودی بدست می آید .

۷.۳. روش رای گیری اکثریت

در ادغام تصمیم گیری ، نتایج دریافتی از طبقه بندی کننده مختلف و واژگان ترکیب می شود و تصمیم نهایی تعیین می شود. ساده ترین راه برای ادغام تصمیم گیری استفاده از رویکرد رأی گیری و انتخاب اینکه کدام یک از برچسب ها بیشترین رای را داشته باشد ، است . رأی گیری یک روش تجمع شناخته شده است که اپراتور رأی دهنده عملگر تو در تو است که از رأی اکثریت برای نتیجه بهبود در پیش بینی زبان آموزان درونی استفاده می کند. برای طبقه بندی ، روش حداکثر احتمال اعمال می شود :

$$X \in w_i \text{ if } g_i(X) > g_j(X), \forall j \neq i, (1)$$

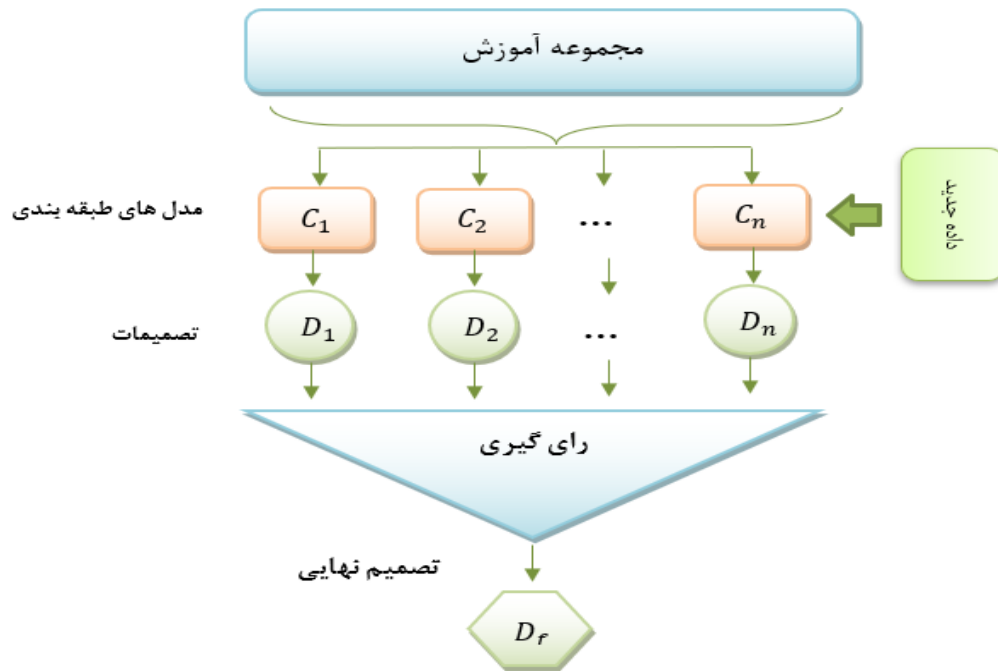
که در آن ، $g_i(X)$ یک تابع طبقه بندی کننده است و w_i کلاس داده انتخاب شده است. در روش MVF ، با توجه به معیار زیر X پیکسل به ω_j اختصاص داده شده است:

$$X \in w_i \text{ if } V(j) = \max_{k=1}^m [V(k)], (2)$$

Where :

$$V(k) = \sum_{i=1}^n \Delta_{ki} (3)$$

$$\Delta_{ki} = \begin{cases} 1 & \text{if } g_k^i(x_i) = \max_{j=1}^n [g_j^i(x_i)] \\ 0 & \text{else} \end{cases} (4)$$



شکل ۳- الگوریتم رای گیری اکثریت

۴. بررسی نتایج

مجموعه داده ای که برای این مقاله استفاده شده است شامل تعداد زیادی نظر کاربر در مورد رفتار و عملکرد پلیس از گروه های مختلف شبکه های اجتماعی می باشد. تعداد نظرات ۵۰۰۵ با کلمات احساسی در هشت کلاس شامل موارد زیر می باشد :

جدول ۱- تعداد نظرات در مجموعه داده پلیس

تعداد احساس در مجموعه داده	احساس
۱۱۸۷	اعتماد
۱۱۴۸	غم
۹۳۶	شادی
۶۱۰	خشم
۴۱۰	تعجب
۳۸۵	انزجار
۱۸۱	ترس
۱۴۷	پیش بینی

۱.۴. معیارهای ارزیابی

در مواردی که دادگان برچسب دار یا حاشیه نویسی شده از نظرات موجود است، چهار معیار درستی، دقت، بازخوانی، ضریب برای ارزیابی کارایی الگوریتم‌ها مطرح است. درستی یعنی درصد کل مشاهداتی است که به درستی طبقه بندی شده‌اند. درستی به صورت زیر محاسبه می‌شود:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (۵)$$

TP شامل شناسایی‌های درست، TN شامل حذف‌های درست، FP شناسایی نادرست و FN حذف‌های نادرست است. به نسبت بین تمام مشاهداتی که به درستی طبقه بندی شده‌اند (TP) به تمام مشاهدات طبقه بندی شده مثبت (FP+TP)، دقت گفته می‌شود.

$$Precision = \frac{TP}{TP+FP} \quad (۶)$$

بازخوانی، نسبت بین مشاهدات به درستی طبقه بندی شده به تمام مشاهدات مثبت است. به این معیار، حساسیت و یا نرخ تأیید درست هم گفته می‌شود.

$$Recall = \frac{TP}{TP+FN} \quad (۷)$$

ضریب f در واقع میانگین هارمونیک بین بازخوانی و دقت است. برای ارزیابی نتایج مدل میتوان از این ضریب (score-f) استفاده کرد و برای محاسبه ی آن از دو پارامتر دقت و بازخوانی استفاده می‌شود.

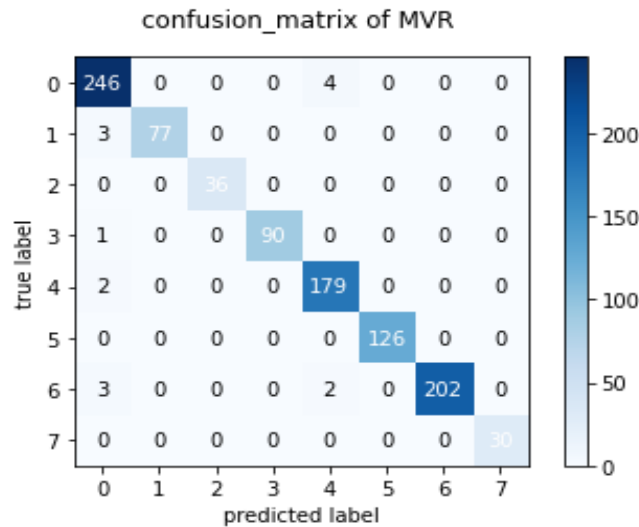
$$f - score = 2 * \frac{Precision \cdot Recall}{Precision + Recall} \quad (۸)$$

بررسی ویژگی‌های در متن فارسی و ارزیابی تأثیر هر گروه از ویژگی‌ها نیازمند تحقیقات متعددی است و در این تحقیق با توجه به اهداف آن مجموعه ای از این ویژگی‌های متنی مورد استفاده در (جدول ۲) در نظر گرفته شده است.

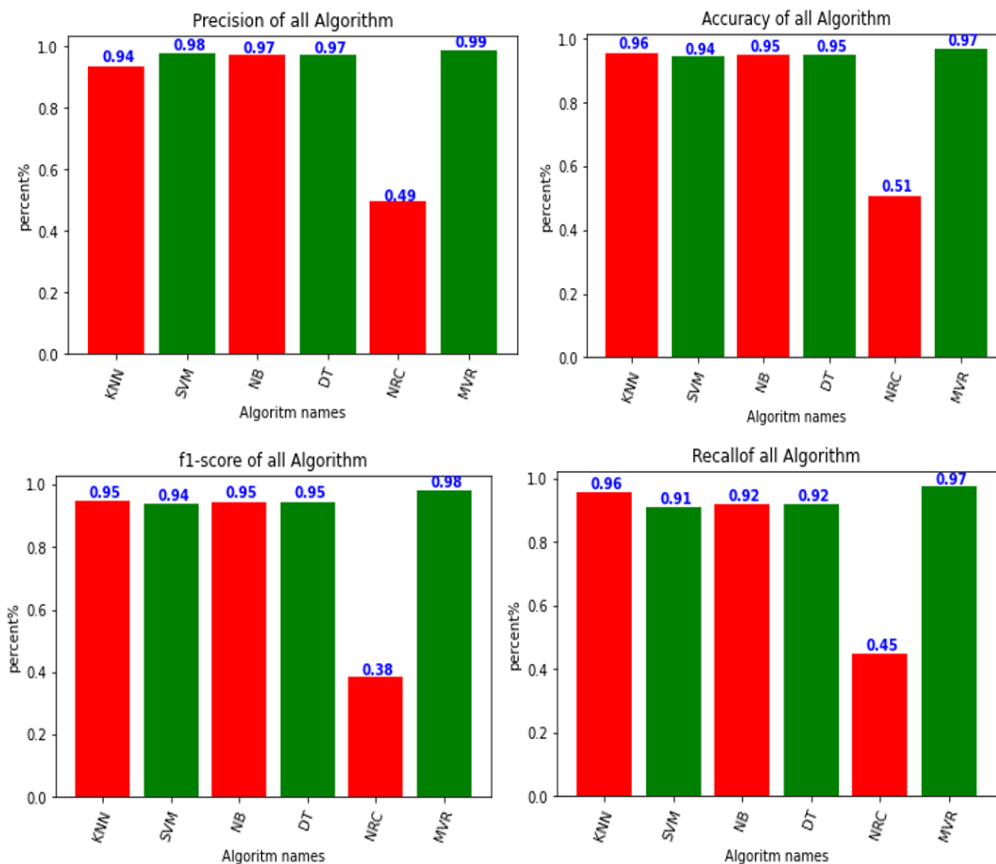
جدول ۲ - لیست ویژگی‌های مورد استفاده در رای گیری اکثریت

ادغام تصمیم گیری	ویژگی	ضریب f	بازخوانی	دقت	درستی
رای گیری اکثریت	Unigram	۰,۹۳	۰,۹۲	۰,۹۴	۰,۹۲
	Bigram	۰,۵۳	۰,۴۷	۰,۷۵	۰,۵۵
	TFIDF	۰,۹۳	۰,۹۱	۰,۹۵	۰,۹۳
	Text To Sequences	۰,۹۸	۰,۹۷	۰,۹۹	۰,۹۷

برای گسترش مدل رای گیری اکثریت ، ماتریس سردرگمی به دست آمده از نتایج طبقه بندی در نظر گرفته می شود .



شکل ۶ - ماتریس سردرگمی رای گیری اکثریت



شکل ۷- مقایسه صحت ارزیابی رای گیری اکثریت با رویکرد ترکیبی

۵. نتیجه گیری

وضعیت سنجی اجتماعات از بعد ملی و امنیت ملی از مهمترین چالش‌ها و همچنین فرصتهاست. شناخت احساسات جمعی می‌تواند توسط دشمنان کشور به مبنایی برای دخالت‌ها و ایجاد تنش‌های اجتماعی و هدایت جامعه و اجتماعات از راه دور مبتنی بر جنگ شناختی تبدیل شود و از طرفی می‌تواند در صورت شناخت مستمر و علمی توسط کشور به مبنایی برای برنامه‌ریزی‌های راهبردی و جلوگیری از آسیب‌های اجتماعی و فراگیر تبدیل گردد و زمینه‌های شادابی و رفاه و امنیت اجتماعی را از این بعد فراهم سازد. در این مقاله، یکی از روش‌های ادغام تصمیم‌گیری به نام رای‌گیری اکثریت اتخاذ شده است که مقایسه‌ای بین طبقه‌بندی‌کننده‌های KNN، SVM، NB، DT و واژگان NRC در مورد نظرکاوی در سطح جمله برای تشخیص احساسات در هشت کلاس (شادی، عصبانیت، غم، انزجار، تعجب، پیش‌بینی، ترس) انجام شده است. پژوهش صورت گرفته به تحلیل احساسات بر روی زبان فارسی بوده که از نقطه قوت آن بشمار می‌رود. در این تحقیق جهت تقویت این موضوع از هر دو رویکرد یادگیری ماشین و واژگان با هدف افزایش دقت تشخیص احساسات نظرات استفاده شده است. در نهایت نتایج حاصل از دو رویکرد مورد اشاره با استفاده از روش‌های ادغام تصمیم‌مانند رأی اکثریت ترکیب شدند. روش پیشنهادی با استفاده از مجموعه داده فارسی تهیه شده مبتنی بر روش‌های استاندارد پیاده‌سازی و ارزیابی شده است. نتایج آزمایشات این پژوهش نشان می‌دهد که در مقایسه با روش‌های قبلی اجرا شده بر روی همین مجموعه داده، استفاده از روش پیشنهادی باعث افزایش معیار صحت با دقت ۹۷٪ در تشخیص نظرات احساسی می‌شود. با توجه به پژوهش انجام شده، پیشنهادهایی

برای پژوهش‌های آتی در این زمینه را می‌توان با TFIDF وزن دار و یا از روش

ادغام تصمیم دمپستر شفر ارائه نمود. دیتاستی که دارای ویژگی‌های متنوعی در این مقاله بوده با نتایج تقریباً خوبی ارزیابی شده که با استفاده از روش پیشنهادی می‌توان به آن حجم و تنوع بیشتری داد و الگوریتم را مجدد مورد بازبینی قرار داد.

۶. مراجع

- [1] Bandari, R., Asur, S., Huberman, B.A., 2012. The pulse of news in social media: forecasting popularity. In: Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media. The AAAI Press, pp. 26–33.
- [2] Kwak, H., Lee, C., Park, H., Moon, S., 2010. What is Twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web. ACM, pp. 591–600.
- [3] N. A. Ghani, S. Hamid, I. A. T. Hashem, and E. Ahmed, “Social media big data analytics: A survey,” *Comput. Human Behav.*, vol. 101, no. December, pp. 417–428, 2019, doi: <https://doi.org/10.1016/j.chb.2018.08.039>.
- [4] Li, X., Peng, Q., Sun, Z., Chai, L., Wang, Y., 2017. Predicting social emotions from readers’ perspective. *IEEE Transactions on Affective Computing*, in press.
- [5] Ali Rashidi , Hassan Ghassemian , EXTENDED DEMPSTER-SHAFER THEORY FOR MULTI- SYSTEM/SENSOR DECISION FUSION , Commission IV, Working Group IV/7.
- [6] ALI J. RASHIDI , A NOVEL DECISION FUSION METHOD BASED ON MULTI-SENSOR BEHAVIOR AND ITS APPLICATION FOR NETWORKED TARGET IDENTIFICATION , Vol. 4, No. 3 (2007) 185–192.
- [7] Aditya Joshi , Vaibhav Tripathi , Ravindra Soni, Pushpak Bhattacharyya, Mark James Carman , EmoGram: An Open-Source Time Sequence-Based Emotion Tracker and Its Innovative Applications , Conference Paper February 2016.
- [8] Adil Majeed , Hasan Mujtaba , Mirza Omer Beg , Emotion Detection in Roman Urdu Text using Machine Learning, 2020 35th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW).
- [9] Ricardo Martins , Marco Gomes , Paulo Novais , Pedro Henriques , Hate speech classification in social media using emotional analysis , 2018 7th Brazilian Conference on Intelligent Systems.