یک مدل تشخیص احساسات معنایی مبتنی بر آنتولوژی و آتوماتای یادگیر عمیق سلولی

* هوشنگ صالحی ** رضا قائمی *** مریم خیرآبادی
* گروه مهندسی کامپیوتر، واحد نیشابور، دانشگاه آزاد اسلامی، نیشابور، ایران amirhoushangsalehi@gmail.com
** گروه مهندسی کامپیوتر، واحد قوچان، دانشگاه آزاد اسلامی، قوچان، ایران r.ghaemi@iauq.ac.ir
*** گروه مهندسی کامپیوتر، واحد نیشابور، دانشگاه آزاد اسلامی، نیشابور، ایران maryam.abadi@gmail.com

تاریخ دریافت: ۱۴۰۲/۰۳/۰۸ تاریخ پذیرش: ۱۴۰۲/۰۴/۰۶
صص: ۹۳- ۱۱۸

چکیده

امروزه شبکه‌های اجتماعی و رسانه‌های ارتباطی نقش به‌سزایی را در زندگی روزمره کاربران دارند. کاربران در زمینه‌های مختلف در شبکه‌های اجتماعی اقدام به گفتگو و تبادل اطلاعات می‌نمایند. در جملات و کامنت‌های کاربران احساسات منفی و مثبت در رابطه با اخبار روز، اتفاقات موجود و غیره وجود دارد که تشخیص این احساسات با چالش‌های زیادی مواجه است. تاکنون روش‌های مختلفی مانند یادگیری ماشین، رویکردهای آماری، هوش مصنوعی و غیره به‌منظور تشخیص احساسات مطرح شده است که علی‌رغم کاربردهای فراوانی که داشته‌اند؛ اما هنوز نتوانسته دقت، شفافیت و صحت قابل قبولی داشته باشند. بنابراین در این مقاله، یک مدل نظرکاوی معنایی مبتنی‌بر آنتولوژی با استفاده از آتوماتای یادگیر عمیق سلولی مبتنی‌بر شبکه عصبی عمیق GMDH ارائه شده است. از رویکرد آنتولوژی برای انتخاب ویژگی‌های برجسته مبتنی‌بر قوانین تولید و از آتوماتای یادگیر عمیق سلولی برای طبقه‌بندی احساسات کاربران استفاده می‌شود. نوآوری اصلی این مقاله الگوریتم پیشنهادی آن است که یک روش یادگیری عمیق جهت پردازش تنها یک عبارت توسعه داده شده و سپس با انتقال آن به حوزه آتوماتای سلولی، پردازش موازی و یا توزیع شده آن فراهم می‌شود. در این مقاله، از مجموعه داده‌های مشتریان آمازون، توئیتر، فیس-بوک، اخبار جعلی COVID-19، آمازون و شبکه اخبار جعلی استفاده شده است. با شبیه‌سازی روش پیشنهادی مشاهده گردید که روش پیشنهادی نسبت به سایر روش‌های دیگر به‌طور میانگین ۳٪ بهبود داشته است.

واژه‌های کلیدی: نظر کاوی، تحلیل احساسات، شبکه عصبی عمیق، آتوماتای سلولی، آنتولوژی.

نوع مقاله: علمی

۱- مقدمه

احساسات و نظرات دیگران در زمان تصمیم‌گیری و انتخاب یک گزینه از میان چند گزینه می‌تواند بسیار حیاتی باشد [۳۱]. به‌دلیل تأثیر احساسات و نظرات بر رفتار انسان می‌توان ادعا نمود که دیدگاه‌ها مرکز فعالیت‌های بشر هستند [۳۲]. شبکه‌های اجتماعی مجازی نسل جدیدی از فضای روابط اجتماعی هستند که توانسته‌اند به‌خوبی در زندگی مردم جای باز کنند. با توجه به دامنه نفوذ روز افزون و تأثیرگذاری این شبکه‌ها، مسأله شناخت الگوهای رفتاری کاربران آن‌ها به یکی از موضوعات جذاب تحقیقی بدل شده است. انجمن‌ها، وبلاگ‌ها، شبکه‌های اجتماعی و سرویس‌های اشتراک محتوا به مردم در اشتراک‌گذاری اطلاعات مفید کمک می‌کند. دریافت احساسات عمومی درباره رویدادهای اجتماعی، فعالیت‌های بازاریابی و اولویت‌های محصول، توجه جوامع علمی و جهان تجارت را به خود جلب کرده است [۳۱]. امروزه، اگر کسی بخواهد محصولی را

نویسنده عهده‌دار مکاتبات: رضا قائمی R.ghaemi@iauq.ac.ir

خریداری نماید، دیگر محدود به نظرات خانواده و دوستان نیست؛ زیرا نظرات و بحث‌های کاربران در مورد محصولات مختلف در وب در دسترس است. برای یک سازمان نیز دیگر نیازی به انجام نظرسنجی و گروه کانونی نیست؛ زیرا چنین اطلاعاتی به‌وفور در اینترنت در دسترس است. با این حال، پیدا کردن و ارزیابی نظرات به‌دلیل گستردگی و تنوع نظرات کار ساده ای نیست. هر سایت معمولاً دارای حجم زیادی نظرات متنی است که رمزگشایی از آن‌ها کار ساده ای نمی‌باشد. به‌طور کلی، استخراج و خلاصه‌سازی چنین اطلاعاتی کار ساده ای نبوده [۳۲] و بنابراین، سیستم‌های تحلیل احساسات خودکار مورد نیاز است. درحال حاضر، پذیرش رشد محتوای تولید شده توسط کاربران در وب سایت‌ها و شبکه‌های اجتماعی مانند توئیتر، آمازون و غیره منجر به افزایش قدرت شبکه‌های اجتماعی برای بیان نظرات در مورد خدمات، محصولات و رویدادها شده است. امروزه افراد و سازمان‌ها به‌طور فزاینده‌ای از محتوای این رسانه‌ها برای تصمیم‌گیری استفاده می‌نمایند [۳۳].

یکی از عمده ابزارها برای تحلیل احساسات، آنتولوژی[۱] می‌باشد. هر چند آنتولوژی‌ها هنوز شکل تکامل یافته‌ای ندارند، اما کاربرد آن‌ها در نظام‌های مختلف هنوز جای پیشرفت دارد. آنتولوژی‌ها که یکی از فناوری‌های اصلی وب معنایی[۲] محسوب می‌شوند، از جمله دستاوردهای هوش مصنوعی هستند که علاوه‌بر داشتن نقش کلیدی در تحقق چشم‌انداز وب معنایی، کاربردهای مختلفی نیز در ارتقاء کیفیت بازیابی اطلاعات کلید واژه‌ای داشته‌اند. آنتولوژی‌ها، با تعریف مفاهیم اصلی یک حیطه موضوعی علمی مبادرت به معرفی یک واژگان مشترک می‌کنند که به واسطه آن، تعامل بین نرم‌افزار و کاربر آسان می‌گردد. سپس با تعیین روابط بین مفاهیم امکان استنتاج معنایی و غنی‌سازی رسایی معنایی را هم برای نمایه‌سازی و هم برای پرسش‌های جستجو فراهم می‌آورند. در واقع، وب معنایی براساس آنتولوژی‌ها و فراداده‌هایی، منابع را با استفاده از آن‌ها نمایه‌سازی می‌کند. آنتولوژی‌ها یک نفش مرکزی در وب معنایی ایفا می‌کنند و به‌طور گسترده‌ای نیز در برنامه‌های کاربردی مدیریت دانش ازآن استفاده شده است[۳۴] و [۳۷].

در تحلیل احساسات به وسیله یادگیری ماشین، سه رویکرد اصلی واژگانی، غیرواژگانی و ترکیبی وجود دارد. یادگیری ماشین به‌طور کلی در دو بخش یادگیری ماشین نظارت‌شده و یادگیری ماشین نظارت‌نشده تقسیم‌بندی می‌شود. یادگیری نظارت‌شده شامل طبقه‌بندی آماری تعداد زیادی از موارد دارای برچسب به‌صورت استاندارد است [۳۶]. بسیاری از اوقات، کاربران در حین ثبت نظر برچسب‌گذاری را از طریق دسته‌بندی یا رتبه‌بندی مشخص در سایت‌ها انجام می‌دهند که در نهایت از طریق این داده‌های دارای برچسب، مدلی برای طبقه‌بندی ایجاد می‌شود که برای تشخیص قطبیت داده جدید مورد استفاده قرار می‌گیرد. ماشین بردار پشتیبان[۳]، شبکه عصبی[۴]، آنتروپی[۵] و طبقه‌بندی کننده ساده بیزین[۶]، روش‌هایی نظارت‌شده می باشند؛ چراکه در این روش‌ها داده‌ها و اطلاعات آماری مربوطه در خصوص ویژگی داده‌ها در دسترس هستند. اما روش واژگانی به هیچ نمونه برچسب‌گذاری شده‌ای نیاز ندارد؛ بنابراین رویکردی بدون نظارت تلقی می‌شود. در این روش، واژگان به دو طبقه مثبت و منفی تقسیم می‌شوند و اگر مجموع تعداد اصطلاحات مثبت بیشتر باشد، سند مثبت در نظر گرفته می‌شود [۳۸] و [۳۹]. یادگیری ماشین زیرمجموعه‌ای تحت عنوان یادگیری عمیق دارد که در آن، روش‌های مختلف شبکه عصبی برای طبقه‌بندی الگوها با یکدیگر ترکیب شده‌اند. در شبکه‌های عصبی عمیق، معمولاً چندین لایه به‌کار گرفته می‌شود تا قادر به یادگیری نمایش داده‌های پیچیده، ویژگی‌های سطح بالاتر و طبقه-بندی صحیح یا اندازه‌گیری خصوصیت داده‌ها باشد. امروزه انواع مختلفی از مدل‌های یادگیری عمیق، برای تجزیه و تحلیل احساسات مانند شبکه عصبی عمیق[۷]، شبکه عصبی بازگشتی[۸] و شبکه عصبی پیچشی[۹] استفاده می‌شوند. با توجه به این‌که روش‌های ذکر شده هر یک دارای معایبی

1. Ontology
2 .Semantic Web

3. Support Vector Machine
4. Neural Network
5. Entropy
6. Naïve Baysian
7. Deep Neural Network
8. Long-Short Term Memory
9. Convolutional Neural Network

هستند و بزرگترین چالش آن‌ها نداشتن دقت، صحت کافی و خطای بالای طبقه‌بندی در تحلیل احساسات است، بنابراین در این مقاله، برای مرتفع نمودن چالش‌های موجود در طبقه‌بندی احساسات از روش آتوماتای یادگیر عمیق سلولی استفاده شده است.

در این مقاله، از ترکیب آنتولوژی و آتوماتای یادگیر عمیق سلولی مبتنی‌بر شبکه عصبی عمیق[1] GMDH[2] برای تحلیل احساسات کاربران در شبکه‌های اجتماعی استفاده می‌گردد. رویکرد آنتولوژی و یادگیری معنایی در مرحله پیش‌پردازش داده‌ها استفاده شده و موجب می‌گردد که جملات ورودی مورد پردازش قرار گرفته و از نظر معنایی آنالیز شوند. هر جمله به چندین کلمه اصلی و هر کلمه به‌صورت معناداری طبقه‌بندی اولیه می‌شوند. سپس با کمک آتوماتای یادگیر سلولی عمیق[3] (D-CLA) اقدام به طبقه‌بندی جملات، کامنت‌ها و توئیت‌های کاربران می‌گردد. از جمله مهم‌ترین مزیت‌های وجود آنتولوژی در مدل پیشنهادی، تقسیم‌بندی مقادیر داده‌ها به‌صورت معنادار به‌منظور انتخاب ویژگی‌های برجسته و بهبود دقت پردازش و تحلیل احساسات کاربران می‌باشد. به‌عنوان نمونه، فرض شود که در توئیت و جملات ارسال‌شده توسط کاربران شبکه اجتماعی، کلماتی منفی وجود دارد. کلمات منفی را از لحاظ معنایی می‌توان به سه دسته بد، خیلی بد و فوق‌العاده بد تقسیم‌بندی نمود. از این‌رو در این مقاله، با به‌کارگیری رویکرد آنتولوژی اقدام به پیش‌پردازش داده‌ها نموده و ویژگی‌ها از نظر معنایی تفکیک می‌گردند. پس از تفکیک معنایی داده‌ها، آتوماتای سلولی و یادگیری عمیق اعمال می‌شود.

اتوماتای سلولی یک مدل ریاضی برای نمایش سیستم‌هاست که در آن اشیایی به‌نام سلول در کنار یکدیگر رفتار سیستم را مدل‌سازی می‌کنند که این مدل‌ها می‌توانند در قالب‌های یک یا چندبعدی تعریف شوند. ساختار همگن و موازی اتوماتای سلولی آن را برای مدل‌سازی انواع مختلف سیستم‌های فیزیکی مناسب ساخته است. برای مدل‌سازی بهینه یک سیستم فیزیکی از ساختار ساده آتوماتای سلولی به‌صورت تعاملات محدود و محلی بین سلول‌ها استفاده می‌شود. ازاین‌رو، از تکنیک آتوماتای سلولی برای بهبود یادگیری عمیق استفاده شده تا بتوان فرآیند طبقه‌بندی و تحلیل احساسات را بهینه نموده و دقت، صحت و سرعت را افزایش داد.

۱-۱- اهداف و نوآوری‌های مقاله

ازجمله مهم‌ترین جنبه‌های جدید بودن و نوآوری این مقاله به شرح ذیل است:

- به‌کارگیری آتوماتای سلولی در هسته مدل شبکه عصبی عمیق به‌عنوان طبقه‌بند نظرات، اخبار و غیره.
- استفاده از آنتولوژی جهت تبدیل و انتخاب ویژگی‌های برجسته براساس مقادیر معنایی.
- آتوماتای یادگیر عمیق سلولی مبتنی‌بر شبکه عصبی عمیق GMDH جهت افزایش دقت تحلیل احساسات.
- ترکیب آنتولوژی و آتوماتای یادگیر عمیق سلولی در سیستم تحلیل احساسات پیشنهادی.

تاکنون روش‌های مختلفی به‌منظور بهبود فرآیند تحلیل احساسات مطرح شده است که علی‌رغم کاربردهای فراوانی که در زمینه بهبود دقت در این زمینه داشته است، اما هنوز با مشکلاتی نظیر عدم داشتن دقت و صحت طبقه‌بندی نظرات و عدم توسعه‌پذیری مواجه می‌باشد. از این‌رو در این مقاله، یک مدل نظرکاوی معنایی مبتنی‌بر آنتولوژی با استفاده از آتوماتای یادگیر عمیق سلولی مبتنی‌بر شبکه عصبی عمیق GMDH جهت افزایش دقت تحلیل احساسات ارائه شده است. در ادامه این مقاله، در بخش ۲ کارهای انجام شده در گذشته مورد بررسی قرار گرفته و در بخش ۳ مدل پیشنهادی و معماری مطرح شده توصیف می‌گردد. در بخش ۴ نتایج بدست‌آمده ارزیابی شده و در بخش ۵ نتیجه‌گیری نهایی مطرح شده است.

۲- پیشینه تحقیق

یو و همکاران (۲۰۱۸)، در مقاله‌ای سیستم تحلیل و پیش‌بینی احساسات مبتنی‌بر محتوای رسانه‌های اجتماعی را بررسی نمودند. در این مقاله، Polaris را پیشنهاد نمودند، سیستمی برای تجزیه و تحلیل و پیش‌بینی مسیرهای احساسی کاربران برای رویدادهایی که در زمان واقعی از محتوای عظیم رسانه‌های اجتماعی تحلیل می‌شوند، و نتایج کار اعتبارسنجی اولیه‌ای را که انجام گرفت، نشان می‌دهد. تحلیل مسیر و تحلیل احساسات را

1. Deep Neural Network
2. Group Model of Handing
3. Deep Cellular Learning Automata

نشان دادند تا کاربران بتوانند در یک نگاه به بینش دست یابند. به‌علاوه، با استفاده از تکنیک یادگیری عمیق، دقت تحلیل و پیش‌بینی احساسات افزایش یافت [۱].

لیو(۲۰۲۰)، در مقاله‌ای تحلیل احساسات متنی براساس مدل CBOW[1] و یادگیری عمیق در محیط کلان داده را بررسی نمود. ابتدا، یک نمایش برداری از متن توسط یک مدل زبان CBOW براساس شبکه‌های عصبی ساخته می‌شود. سپس، شبکه عصبی پیچشی از طریق مجموعه آموزشی برچسب‌گذاری شده آموزش داده می‌شود تا ویژگی‌های معنایی متن را به تصویر بکشد. نتایج تجربی بر روی مجموعه داده‌های COAE2014 و IMDB نشان می‌دهد که این روش می‌تواند به‌طور دقیق دقت دسته‌بندی احساسی متن را در دو مجموعه داده به‌ترتیب با مقادیر ۹۰,۵٪ و ۸۷,۲٪ تعیین کند [۲].

سینگ و همکاران (۲۰۲۰)، در مقاله‌ای به بررسی و تحلیل مقایسه‌ای تحلیل احساسات در رسانه‌های اجتماعی پرداختند. این مقاله مروری جامع از تکنیک تحلیل احساسات براساس تحقیقات اخیر ارائه می‌کند و متعاقباً یادگیری ماشین (ماشین بردار پشتیبان، نیوبیزین، رگرسیون خطی و جنگل تصادفی[2]) و تکنیک‌های استخراج ویژگی (اجماع اثبات سهام یا POS[3]، BOW و برچسب‌گذاری HASS) را در زمینه تحلیل احساسات در رسانه‌های اجتماعی بررسی می‌کند. مجموعه داده‌های توئیتر بیشتر با چارچوب پیشنهادی مورد بررسی و پیش-پردازش قرار می‌گیرند که حقایق متقاطع در مورد قابلیت‌ها و کمبود روش‌های تحلیل احساسات را به‌دست می‌دهد. POS مناسب‌ترین تکنیک استخراج ویژگی با ماشین بردار پشتیبان و طبقه‌بندی کننده نیوبیزین است، در حالی که جنگل تصادفی و رگرسیون خطی نتایج بهتری را با برچسب‌گذاری HASS ارائه می‌دهند [۳].

ماندلوی و همکاران (۲۰۲۰)، در مقاله‌ای به تجزیه و تحلیل احساسات توئیتر با استفاده از روش‌های یادگیری ماشینی پرداختند. در این مقاله، روش‌های مختلف یادگیری ماشین مانند طبقه‌بندی ساده بیز، طبقه‌بندی ماشین بردار پشتیبان و طبقه‌بندی حداکثر آنتروپی با یکدیگر مقایسه گردید. تجزیه و تحلیل داده‌های توئیتر در جنبه‌های مختلف برای استخراج احساسات انجام می‌شود. این مطالعه نشان می‌دهد که روش یادگیری ماشینی مانند نیوبیزین دارای بالاترین دقت است و می‌تواند به-عنوان روش‌های یادگیری پایه و همچنین در برخی موارد روش‌های حداکثر آنتروپی بسیار موثر در نظر گرفته شود [۴].

چن و همکاران (۲۰۲۰)، در مقاله‌ای به کاوش در رسانه-های اجتماعی برای تجزیه و تحلیل احساسات با استفاده از یادگیری عمیق پرداختند. این مطالعه از کاربرد Militarylife PTT بزرگترین انجمن آنلاین تایوان به-عنوان منبع داده‌های تجربی خود استفاده کرد. هدف از این مطالعه، ساخت چارچوب و فرآیندهای تحلیل احساسات برای رسانه‌های اجتماعی به‌منظور پیشنهاد فرهنگ لغت احساسات نظامی خود توسعه‌یافته برای بهبود طبقه‌بندی احساسات و تحلیل عملکرد مدل‌های مختلف یادگیری عمیق با ترکیب‌های کالیبراسیون پارامترهای مختلف بود. نتایج تجربی نشان می‌دهد که دقت و اندازه‌گیری F1 مدلی که فرهنگ لغت‌نامه‌های احساسات موجود و فرهنگ لغت احساسات نظامی خود توسعه‌یافته را ترکیب می‌کند، بهتر از نتایج حاصل از استفاده از واژه‌نامه‌های احساسات موجود است. علاوه‌بر این، دقت و اندازه‌گیری F1 عملکرد بهتری برای طبقه‌بندی احساسات دارند [۵].

چاوهان و همکاران (۲۰۲۰)، در مقاله‌ای به تجزیه و تحلیل جامع انواع قید برای استخراج احساسات کاربر در بررسی محصولات آمازون پرداختند. برای طبقه‌بندی احساسات، عنصر اصلی بهره‌برداری از کلمات حامل قطبی است که در بررسی‌ها وجود دارد، به‌عنوان مثال، صفات، افعال و قیدها و غیره. مطالعات مختلف اهمیت اشکال مختلف قیدها را در کار طبقه‌بندی احساسات نشان می‌دهد. در ادبیات، گزارش شده است که قیدهای عمومی به‌شدت به طبقه‌بندی احساسات با دقت بهتر کمک می‌کنند، در حالی که بقیه نشان می‌دهند که قیدهای درجه برای طبقه‌بندی احساسات مهم هستند. در این مقاله تأثیر قیود بر روی دو محصول، محصولات اداری و DVDهای موسیقی خریداری شده از آمازون ارزیابی شده است. نتایج مطالعه نشان داد که دو قید فوق‌العاده کلی و

1. Continuous Bag of Words
2 .Random Forest
3 .Proof of Stack

قید درجه-WH تأثیر بیشتری نسبت‌به سایر اشکال قید دارند [۶] .

بهرا و همکاران (۲۰۲۱)، در مقاله‌ای مدل شبکه عصبی بازگشتی برای تجزیه و تحلیل احساسات در داده‌های بزرگ اجتماعی ارائه نمودند. در این مقاله، یک رویکرد ترکیبی از دو معماری یادگیری عمیق، یعنی شبکه عصبی پیچشی و حافظه کوتاه‌مدت-بلندمدت برای طبقه‌بندی احساسات نظرات ارسال‌شده در حوزه‌های مختلف پیشنهاد شده‌است. شبکه‌های کانولوشنال عمیق در انتخاب ویژگی محلی بسیار مؤثر بوده‌اند، در حالی که شبکه‌های تکراری حافظه کوتاه‌مدت-بلندمدت اغلب نتایج خوبی در تجزیه و تحلیل متوالی یک متن طولانی دارند. این آزمایش بر روی چهار مجموعه داده مروری از حوزه‌های مختلف انجام شده است تا مدلی را آموزش دهد که می‌تواند انواع وابستگی‌هایی را که معمولاً در یک پست ایجاد می‌شود، مدیریت کند. نتایج تجربی نشان می‌دهد که مدل مجموعه پیشنهادی از نظر دقت و سایر پارامترها بهتر از سایر رویکردهای یادگیری ماشین عمل می‌کند [۷].

چاندار و همکاران (۲۰۲۱)، در مقاله‌ای تحلیل معنایی احساسات از طریق مجموعه توئیتر استخراج‌شده از وب را بررسی نمودند. در این مقاله، در مورد تجزیه و تحلیل احساسات در یک مجموعه داده توئیتر با توئیت‌های مختلف از کاربران مختلف بررسی گردید. بنابراین، از تکنیک‌های طبقه‌بندی مختلف مانند ماشین بردار پشتیبان، رگرسیون لجستیک، رگرسیون لجستیک با بهینه‌ساز گرادیان تصادفی، طبقه‌بندی درخت تصمیم، حافظه کوتاه‌مدت-بلندمدت دو جهته و طبقه‌بندی جنگل تصادفی برای تجزیه و تحلیل احساسات افراد استفاده شده است. نتایج نشان داد که با مشاهده منحنی ROC و امتیاز دقت مشخص شد که حافظه کوتاه‌مدت-بلندمدت دوطرفه با دقت ۷۸/۹۰٪ بهترین طبقه‌بندی کننده است و در یافتن تحلیل احساسات بسیار مفید است [۸].

پاتاک و همکاران (۲۰۲۱)، در مقاله‌ای به تجزیه و تحلیل احساسات کاربران در شبکه‌های اجتماعی پرداختند. در این مقاله، یک مدل تحلیل احساسات در سطح موضوعی مبتنی‌بر یادگیری عمیق پیشنهاد گردید. تازگی رویکرد پیشنهادی این است که در سطح جمله کار می‌کند تا موضوع را با استفاده از نمایه‌سازی معنایی پنهان آنلاین با محدودیت منظم‌سازی استخراج کند و سپس، مکانیسم توجه سطح موضوع را در شبکه حافظه کوتاه‌مدت برای انجام تحلیل احساسات اعمال می‌کند. برای مجموعه داده‌های SemEval- Task 4 Subtask B 2017 به‌عنوان موردی از تجزیه و تحلیل احساسات در سطح موضوع درون دامنه، میانگین یادآوری ۰,۸۷۹ به‌دست آمده است، در حالی که برای داده‌های خارج از دامنه، میانگین یادآوری ۰,۸۴۶، ۰,۸۲۴ و ۰,۷۹۴ حاصل شده است [۹].

آوجان و همکاران (۲۰۲۱)، در مقاله‌ای تکنیک تجزیه و تحلیل احساسات و نظریه مجموعه‌های نوتروسوفیک (NS[1]) برای استخراج و رتبه‌بندی کلان داده‌ها از بررسی‌های آنلاین ارائه نمودند. این مقاله تحقیقاتی، روش جدیدی را معرفی می‌کند که نظریه مجموعه‌های نوتروسوفیک را در تکنیک تجزیه و تحلیل احساسات و تصمیم‌گیری چند ویژگی (MADM[2]) ادغام می‌کند تا محصولات مختلف را براساس بررسی‌های آنلاین متعدد رتبه‌بندی کند. این روش از دو بخش تشکیل شده است، شامل تعیین امتیازات احساسات بررسی‌های آنلاین براساس تکنیک تجزیه و تحلیل احساسات و رتبه‌بندی محصولات جایگزین از طریق نظریه مجموعه‌های نوتروسوفیک. یک مطالعه موردی با مجموعه داده‌های واقعی توئیتر برای نشان‌دادن کاربرد روش پیشنهادی ارائه شده است. نتایج عملکرد خوبی را در مدیریت داده‌های خنثی در مرحله تعیین امتیازات احساسات و همچنین مرحله رتبه‌بندی نشان می‌دهد [۱۰].

کای و همکاران (۲۰۲۲)، در مقاله‌ای یک مدل توصیه عمیق از احساسات متقابل نظرات و رتبه‌بندی‌های کاربران ارائه نمودند. این مقاله یک مدل توصیه یادگیری عمیق به‌عنوان نمونه DeepCGSR[3] را پیشنهاد می‌کند که احساسات مرور متنی و ماتریس رتبه‌بندی را ادغام می‌کند. آزمایش‌های تکراری روی مجموعه داده‌های تجارت الکترونیک از آمازون نشان می‌دهد که DeepCGSR به‌طور مداوم از مدل‌های توصیه‌شده LFM، SVD++، DeepCoNN، TOPICMF و

1. Neutrosophy Sets Theory
2. Multiple Attribute Decision Making
3. Deep Cross-Grained Sentiments Rating

NARRE بهتر عمل می‌کند. DeepCGSR که براساس مطالعات و یافته‌های قبلی ساخته شده است، پیشرفته-ترین حالت است که طراحی و توسعه الگوریتم‌های توصیه را با دقت توصیه بهبودیافته به‌جلو می‌برد [۱۱].

رواتی و همکاران (۲۰۲۲)، در مقاله‌ای به تجزیه و تحلیل احساسات با استفاده از یادگیری ماشین پرداختند. استخراج احساسات و طبقه‌بندی با چندین رویکرد از جمله الگوریتم‌های عصبی فازی و بهینه‌سازی انجام می‌شود. در این مقاله که از شبکه عصبی دوگانه پیشرو استفاده شده است، هنگامی‌که داده‌های بلادرنگ حاوی کاراکترها و جریان اطلاعات متعددی باشد، این رویکردها در طبقه‌بندی بی‌اثر هستند. برای دستیابی به طبقه‌بندی ماهرانه، از شبکه عصبی پیشرو دوگانه استفاده می‌شود و اطلاعات لایه خروجی به لایه دوگانه شبکه منتقل می‌شود. ازاین‌رو، اطلاعات بهینه‌شده و به‌طور موثر پردازش می‌شوند، که در نتیجه طبقه بندی احساسات به‌دست می‌آید. کل فرآیند الگوریتم انجام می‌شود و نتایج به‌دست آمده با الگوریتم عصبی فازی و بهینه‌سازی مقایسه می‌شود. DFFNN[1] از نظر پارامترهای طبقه‌بندی از الگوریتم موجود بهتر عمل می‌کند [۱۲].

بیرادار و همکاران (۲۰۲۲)، در مقاله‌ای از یادگیری ماشین برای کاوش تجزیه و تحلیل احساسات در داده‌های توئیتر استفاده نمودند. هدف این مقاله، توسعه فناوری کلان داده مورد استفاده برای جمع‌آوری و مدیریت داده‌های بزرگ بدون ساختار از رسانه‌های اجتماعی بلادرنگ برای تحلیل احساسات برای شناسایی برند و خدمات بود. این روش یک الگوریتم مبتنی‌بر تجزیه و تحلیل احساسات با استفاده از طبقه‌بندی مرور مشتریان ابداع کرد که با در اختیار داشتن مجموعه داده‌ها، خوشه‌بندی داده‌ها براساس حوزه‌های خاص، بردار ویژگی با استفاده از مدل‌های N-Gram و بردارهای TF-IDF[2] مترادف‌ها و تحلیل احساسات طبقه‌بندی سروکار داشت. نتیجه نشان می‌دهد که تحلیل احساساتی کاربردی با خوشه‌بندی بدون نظارت داده‌ها در حوزه‌های خاص و تکنیک‌های یادگیری ماشینی نظارت‌شده، حجم زیادی از داده‌های توئیتر را به روشی کارآمد مدیریت می‌کند [۱۳].

ویلگاس و همکاران (۲۰۲۲)، در مقاله‌ای روشی برای تحلیل احساسات در شبکه‌های اجتماعی در محیط R پیشنهاد نمودند. در این مقاله، هدف تعیین احساسات مردم نسبت‌به یک برند، یک محصول یا یک خدمات و حتی شناسایی واکنش افراد به رویدادها و روندهای ایجاد شده در محیط آن‌ها است. این کار روشی را برای تجزیه و تحلیل احساسات در شبکه‌های اجتماعی برای برقراری ارتباط داده‌های عظیم در مورد هر موضوع مورد علاقه پیشنهاد می‌کند. برای ارزیابی روش، یک مطالعه موردی که به دنبال شناسایی احساسات موجود جمعیت در مورد مدیریت دولت اکوادور است، اجرا شد. با استفاده از یک روش قوی در مجموعه توئیت‌ها، می‌توان احساسات را به اندازه کافی شناسایی کرد. برای رسیدن به این هدف، از ابزارهای قوی مانند R Studio و Power BI برای تجسم نتایج استفاده شد. نتایج نشان داد که حجم زیاد داده‌ها در شبکه‌های اجتماعی امکان اجرای فرآیندهای تحلیلی مختلف را فراهم می‌کند و به تصمیم‌گیری سازمان‌ها کمک می‌کند [۱۴].

جین و همکاران (۲۰۲۲)، مقاله‌ای با عنوان محاسبات الهام‌گرفته از شناخت هوشمند با چارچوب تجزیه و تحلیل داده‌های بزرگ برای تجزیه و تحلیل احساسات و طبقه‌بندی ارائه نمودند. این مقاله، محاسبات شناختی جدیدی را با ابزار تجزیه و تحلیل کلان داده برای تجزیه و تحلیل احساسات ارائه می‌کند. مدل پیشنهادی شامل فرآیندهای مختلفی مانند پیش‌پردازش، استخراج ویژگی، انتخاب ویژگی و طبقه‌بندی است. برای مدیریت کلان داده ها از ابزار Hadoop Map Reduce استفاده می‌شود. علاوه‌بر این، یک الگوریتم بهینه‌سازی طوفان مغزی باینری (BBSO[3]) برای فرآیند انتخاب ویژگی و در نتیجه دستیابی به عملکرد طبقه‌بندی بهبودیافته استفاده می‌شود. تجزیه و تحلیل نتایج تجربی جامع، عملکرد بهتر مدل BBSO-FCM ارائه شده در مجموعه داده معیار را تضمین می‌کند [۱۵].

مایتی و همکاران (۲۰۲۳)، جهت تجزیه و تحلیل احساسات متن از شبکه های عصبی کانولوشنال چند کانالی و ساختار نحوی استفاده نمودند. نوع الگوریتم شبکه عصبی عمیق استفاده شده در تحقیق آنها الگوریتم

1. A Deep Feed Forward Neural Network
2. Term Frequency-Inverse Document Frequency
3. Binary Brain Storm Optimization

شبکه عصبی عمیق CNN و LSTM چندکاناله بوده است. ابزار استفاده شده جهت شبیه‌سازی متد ترکیبی پیشنهادی زبان برنامه‌نویسی متلب می‌باشد. دقت تشخیص روش آنها جهت تشخیص احساسات ۸۷,۷۵٪ بوده که نسبت به سایر روشهای مستقل همچون LSTM, CNN بهبود قابل ملاحظه‌ای داشته است [۴۰].

ایرون و همکاران (۲۰۲۳)، جهت تحلیل احساسات بر روی مجموعه داده‌های به زبان رومانیایی از روش RST[1] استفاده نمودند. زبان برنامه نویسی جهت تشخیص احساسات متون رومانیایی پایتون می‌باشد. نتایج دقت تحقیق آنها جهت تشخیص احساسات ۸۶٪ بوده است [۴۱].

یونائیزا و همکاران (۲۰۲۳)، جهت تشخیص احساسات در خصوص ویروس آمیکرون مرتبط به بیماری کرونا (COVID-19) در رسانه اجتماعی توئیتر از روشهای یادگیری ماشین همچون نایوبیز، رگرسیون خطی، ماشین بردار پشتیبان و درخت تصمیم استفاده نمودند. آنها جهت تشخیص احساسات و آنالیز دیتاست توئیتر از ابزار برنامه نویسی متلب استفاده نمودند. دقت تشخیص روش مطرح شده در تحقیق آنها ۸۷٪ بوده است [۴۲].

جدول (۱)، مقایسه‌ای از سوابق پیشین جهت تحلیل و تشخیص احساسات را نشان می‌دهد.

1. Romanized Sindhi Text

جدول ۱. مقایسه سوابق پیشین جهت تحلیل و تشخیص احساسات

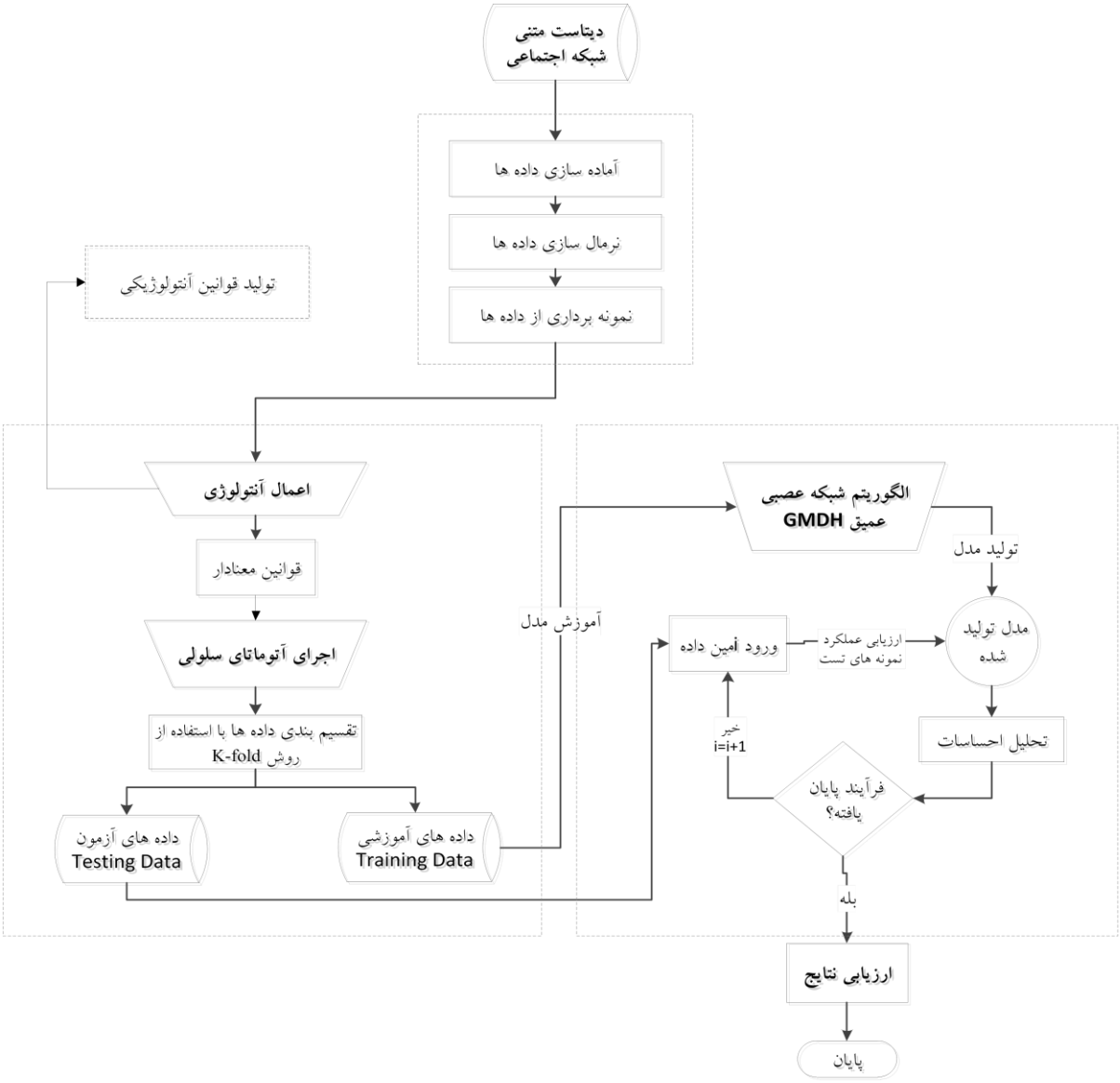
نویسندگان/ سال	روش پیشنهادی	مزایای روش پیشنهادی	معایب روش پیشنهادی	مجموعه داده آزمون
یو و همکاران/۲۰۱۸	تجزیه و تحلیل و پیش‌بینی مسیرهای احساسی کاربران	*دقت و صحت مطلوب*	زمان‌بر بودن پردازش، عدم مقیاس‌پذیری	مجموعه داده‌های توئیتر
لیو/۲۰۲۰	تحلیل احساسات متنی براساس مدل CBOW و یادگیری عمیق در محیط کلان داده	دقت و صحت مطلوب	زمان‌بر بودن پردازش، عدم مقیاس‌پذیری	مجموعه داده‌های COAE2014 و IMDB
سینگ و همکاران/۲۰۲۰	تحلیل احساسات با استفاده از روش‌های یادگیری ماشین و تکنیک‌های استخراج ویژگی	زمان پردازش سریع	ضعف دقت و صحت تحلیل احساسات	مجموعه داده‌های توئیتر
ماندلوی و همکاران/۲۰۲۰	تجزیه و تحلیل احساسات توئیتر با استفاده از روش‌های یادگیری ماشین	زمان پردازش سریع	عدم مقیاس‌پذیری	مجموعه داده‌های توئیتر
چن و همکاران/۲۰۲۰	کاوش در رسانه‌های اجتماعی برای تجزیه و تحلیل احساسات با استفاده از یادگیری عمیق	دقت مطلوب	زمان‌بر بودن پردازش	مجموعه داده‌های انجمن آنلاین تایوان
چاوهان و همکاران/۲۰۲۰	تجزیه و تحلیل جامع انواع قید برای استخراج احساسات کاربر در بررسی محصولات آمازون	دقت و صحت مطلوب	زمان‌بر بودن پردازش، عدم مقیاس‌پذیری	مجموعه داده‌های آمازون
بهرا و همکاران/۲۰۲۱	مدل کانولوشنال LSTM برای تجزیه و تحلیل احساسات در داده-های بزرگ اجتماعی	مرتب‌سازی داده‌ها در حجم زیاد	عدم تحلیل مجموعه داده‌های چندزبانه	چهار مجموعه داده از حوزه‌های متفاوت
چاندار و همکاران/۲۰۲۱	تحلیل احساسات با استفاده از تکنیک‌های طبقه‌بندی	زمان پردازش سریع و مقیاس‌پذیری بالا	ضعف دقت و صحت تحلیل احساسات	مجموعه داده‌های توئیتر
پاتاک و همکاران/۲۰۲۱	تجزیه و تحلیل احساسات کاربران با استفاده از روش یادگیری عمیق	دقت و صحت مطلوب	زمان‌بر بودن پردازش، عدم مقیاس‌پذیری	مجموعه داده‌های SemEval- 2017 Task 4 Subtask B
آوجان و همکاران/۲۰۲۱	تجزیه و تحلیل احساسات و نظریه مجموعه‌های نوتروسوفیک برای استخراج و رتبه‌بندی کلان داده‌ها	تجزیه و تحلیل به‌موقع	ضعف دقت و صحت تحلیل احساسات	مجموعه داده‌های توئیتر
کای و همکاران/۲۰۲۲	تحلیل احساسات با یادگیری عمیق	زمان پردازش سریع و مقیاس‌پذیری بالا	ضعف دقت و صحت تحلیل احساسات	مجموعه داده‌های تجارت الکترونیک از آمازون
رواتی و همکاران/۲۰۲۲	تجزیه و تحلیل احساسات با استفاده از یادگیری ماشین	زمان پردازش سریع	ضعف دقت و صحت تحلیل احساسات	مجموعه داده‌های توئیتر

بیرادار و همکاران/۲۰۲۲	تحلیل احساسات با استفاده از روش‌های یادگیری ماشین	زمان پردازش سریع و مقیاس‌پذیری بالا	ضعف دقت و صحت تحلیل احساسات	مجموعه داده‌های توئیتر
ویلگاس و همکاران/۲۰۲۲	تحلیل احساسات در شبکه‌های اجتماعی با استفاده از R	دقت و صحت مطلوب	زمان‌بر بودن پردازش، عدم مقیاس‌پذیری	مجموعه داده‌های توئیتر
جین و همکاران/۲۰۲۲	تحلیل احساسات با استفاده از روش کلان داده	زمان پردازش سریع و مقیاس‌پذیری بالا	ضعف دقت و صحت تحلیل احساسات	مجموعه داده‌های توئیتر
مایتی و همکاران/۲۰۲۳	جهت تجزیه و تحلیل احساسات متن از شبکه های عصبی کانولوشنال چند کانالی و ساختار نحوی	زمان پردازش سریع و مقیاس‌پذیری بالا	ضعف دقت و صحت تحلیل احساسات	مجموعه داده‌های توئیتر
ایرون و همکاران/۲۰۲۳	تحلیل احساسات بر روی مجموعه داده‌های به زبان رومانیایی از روش RST	دقت و صحت مطلوب	زمان‌بر بودن پردازش، عدم مقیاس‌پذیری	مجموعه داده‌های توئیتر
یونائیزا و همکاران/۲۰۲۳	تشخیص احساسات در خصوص ویروس آمیکرون مرتبط به بیماری کرونا (COVID-19) در رسانه اجتماعی توئیتر	زمان پردازش سریع و مقیاس‌پذیری بالا	ضعف دقت و صحت تحلیل احساسات	مجموعه داده‌های توئیتر

با توجه به بررسی سوابق پیشین و تحقیقاتی که اخیرا در زمینه نظرکاوی و تحلیل احساسات مطرح شده است، مشاهده گردید که تاکنون، روش‌های مختلفی جهت تحلیل احساسات کاربران در رسانه‌های اجتماعی ارائه گردیده است که علی‌رغم کاربردهای فراوانی که داشته‌اند؛ اما هنوز با چالش‌های عدم داشتن دقت کافی، خطای بالا و زمان‌بر بودن تحلیل احساسات مواجه هستند. ازاین‌رو، از ترکیب رویکرد آنتولوژی و آتوماتای یادگیر عمیق سلولی مبتنی‌بر شبکه عصبی عمیق GMDH جهت تحلیل احساسات کاربران شبکه‌های اجتماعی در ارتباط با مشتریان آمازون، توئیت‌های توئیتر، کامنت‌های فیس‌بوک، اخبار جعلی COVID-19 و خبرهای شبکه اخبار جعلی به‌منظور مرتفع نمودن چالش‌های عدم داشتن دقت کافی، خطای بالا و زمان‌بر بودن تحلیل احساسات استفاده می‌گردد.

۳- روش پیشنهادی

در شکل (۱)، معماری مدل پیشنهادی جهت تحلیل و تشخیص احساسات در شبکه اجتماعی نشان داده شده است.



شکل ۱. معماری مدل پشنهادی جهت تحلیل و تشخیص احساسات در شبکه اجتماعی

در ابتدا، مجموعه داده شبکه‌های اجتماعی به بخش پیش‌پردازش در معماری پیشنهادی وارد می‌شود. سپس فرآیند پیش‌پردازش کلمات انجام‌شده و حروف اضافه حذف می‌گردند. حذف نمونه‌های تکراری نه‌تنها تأثیر منفی در مدل ایجادشده نمی‌گذارد، بلکه باعث کاهش زمان یادگیری مدل نیز شده و باعث می‌شود محاسبات تکراری کاهش یابند. پس از اعمال پیش‌پردازش معمولی بر روی داده‌ها، یک مجموعه داده منسجم[1] ایجاد می‌گردد. سپس مجموعه داده‌های پیش‌پردازش شده به رویکرد آنتولوژی وارد می‌شوند. رویکرد آنتولوژی اقدام به تفکیک کلمات و استخراج وزن کلمات از دیکشنری و پایگاه دانش قوانین تولید شده نموده و براساس میزان مثبت یا منفی بودن، کلمه را درجه‌بندی کرده و درنهایت، برای هر کلمه یک مفهوم معناداری تعیین می‌کند. رویکرد آنتولوژی موجب می‌شود که مدل تولیدشده در مرحله بعد، از دقت بالاتری برخوردار بوده و تحلیل احساسات با سرعت بالاتری صورت پذیرد. بنابراین، در این مقاله، از آنتولوژی جهت تحلیل داده‌ها و استخراج مجموعه‌ای از قوانین معنادار استفاده می‌گردد. سپس از قوانین معنادار جهت انتخاب ویژگی‌های تاثیرگذار استفاده می‌شود.

در مرحله بعدی، داده‌ها به دو دسته آموزشی و آزمایشی تقسیم می‌شوند. با استفاده از روش K-fold، داده‌ها به K قسمت برای ایجاد داده‌های آموزش و آزمایش، تقسیم می‌شوند. داده‌های آموزشی به آتوماتای یادگیر عمیق سلولی مبتنی‌بر شبکه عصبی عمیق GMDH وارد شده و مدل عصبی تولید می‌گردد. سپس داده‌های آزمایش به مدل تولیدشده وارد گردیده و مورد تجزیه و تحلیل قرار می‌گیرد. ارزیابی نهایی براساس داده‌های آزمایش می‌باشد. همان‌طور که در شکل (۱) مشاهده شد، به‌طور کلی فرآیند طبقه‌بندی داده‌های شبکه اجتماعی عبارتند از پیش‌پردازش و تفکیک داده‌ها، آماده‌سازی داده‌ها، نرمال‌سازی کلان‌داده‌ها، آنتولوژی، تفکیک نمونه‌های آموزشی و آزمایشی، اعمال الگوریتم شبکه عصبی سلولی عمیق که شامل آتوماتای یادگیری سلولی و شبکه عصبی عمیق GMDH می‌شود. در ادامه، به تشریح هر یک از این فرآیندها پرداخته می‌شود.

1. Cohesive

۳-۱- پیش‌پردازش داده‌ها

در فرآیند پیش‌پردازش داده‌ها چندین مرحله انجام می‌شود که عبارتند از:

- **حذف نویز**

مجموعه داده‌های استفاده شده در این مقاله دارای نویز یا مقادیر پرت است. در این مقاله، از یک رویکرد ساده جهت حذف نویز در دیتاست استفاده می‌گردد. رویکرد استفاده شده با پیمایش کل داده‌ها اقدام به حذف نمودنه‌هایی می کند که دارای مقادیر پرت می‌باشد. بنابراین، فرآیند حذف نویز به‌منظور از بین‌بردن ویژگی‌هایی که نسبت‌به بقیه ویژگی‌ها پرت است، مورد استفاده قرار می‌گیرد. ویژگی‌هایی که دارای رفتاری متفاوت با سایر ویژگی‌هاست، به‌عنوان ویژگی‌های پرت شناسایی و از مجموعه داده حذف می‌گردد.

- **حذف ایست کلمات**

در مجموعه داده‌ها از برخی ایست کلمات استفاده می‌شود. کلماتی معمول مانند a و an و the به‌عنوان ایست کلمات شناخته می‌شوند.با توجه به این‌که این کلمات هیچ نوع اطلاعات و احساساتی را ارایه نمی‌کنند، از مجموعه داده‌ها حذف می‌شوند. جهت حذف ایست کلمات کافی است کل مجموعه داده مربوطه پیمایش‌شده و با یک فیلتر ساده، ایست کلمات از مجموعه داده حذف شوند.

- **تبدیل کلیه حروف به حروف کوچک**

به‌منظور جلوگیری از اشتباه در حروف کوچک در کلمات، کلیه حروف به حروف کوچک تبدیل می‌شود.

- **تبدیل رشته پیام به بردار ویژگی**

براساس ویژگی‌های غیرقابل پیش‌بینی در نظرات و همچنین برای استاندارد کردن نظرات کاربران که از ورودی گرفته می‌شوند، پیش‌پردازش انجام شود. در ادامه، مرحله پیش‌پردازش کلیه کلمات به‌کار رفته در مجموعه داده مرور شده و به شمارش تعداد تکرار هر کلمه در کل مجموعه داده پرداخته می‌شود. به‌منظور انجام پردازش اولیه بر روی داده‌ها، کلمه‌هایی که تنها یک‌بار در کل مجموعه داده تکرار شده بودند، حذف گردید. همچنین به‌منظور حذف کلمه‌های حاوی حروف اضافه که به تعداد زیادی در کل مجموعه داده تکرار می‌شوند، کلمه‌هایی که تعداد تکرار آن‌ها بیشتر از نیمی از تعداد رکوردها باشد نیز

از مجموعه لغات حذف می‌شوند. بنابراین در مجموعه اولیه انتخاب‌شده کلمات، کلمه‌هایی که فرکانس رخداد کمتر از ۲ و بیش از نیمی از تعداد کل نظرات را شامل می‌شوند، از درون داده حذف می‌گردد.

۳-۲- انتخاب ویژگی با استفاده از رویکرد آنتولوژی

آنتولوژی دانش مشترکی از یک حوزه خاص بین افراد و سیستم‌هاست. این زبان، به زبان خاصی به‌نام زبان آنتولوژی وب (OWL[1]) نوشته شده است. برای دستیابی به کارایی آنتولوژی پیشنهادی، یک آنتولوژی کلاسیک با استفاده از Protégé OWL طراحی می‌گردد [۱۶]. در این مقاله، ازآنتولوژی جهت انتخاب ویژگی‌های برجسته از بردار ویژگی کلمات در شبکه‌های اجتماعی و رسانه‌های ارتباطی همچون فیس‌بوک، توئیتر، آمازون و غیره استفاده می‌گردد. معماری آنتولوژی پیشنهادی مجموعه‌ای از قوانین را تولید نموده که بتوان از این قوانین جهت انتخاب ویژگی‌های برجسته استفاده نمود. آنتولوژی پیشنهادی موجب کاهش ابعاد داده‌ها شده و درنهایت، دقت تشخص احساسات را بهبود می‌بخشد.

آنتولوژی پیشنهادی مجموعه‌ای از قوانین را به‌عنوان پایگاه دانش تولید می‌کند. در این بخش، قوانینی که از آنتولوژی برای تشخیص احساسات کاربران تولید می‌گردد را توضیح می‌دهد. براساس ویژگی‌های خاصی که بین زیرکلاس و صفات وجود دارد، قوانین آنتولوژی تولید می‌شوند. تعداد کل ویژگی‌های شناسایی شده چهار مورد است شامل وزن مثبت یا منفی کلمه، تعداد تکرار کلمه در جمله، تعداد کلمه در کل پایگاه داده و وزن کلمه در دیکشنری. براساس ویژگی‌های به‌دست‌آمده، تعدادی روابط ساخته می‌شود که نشان‌دهنده قوانین مربوط به کلاس‌ها، زیرکلاس‌ها و نمونه‌ها هستند. برخی از روابط شکل گرفته بین زیرکلاس‌ها ونوع احساسات کاربران اعم از مثبت یا منفی بودن بدین صورت است که کلمه‌ای که دارای وزن منفی است، احتمال دارن جمله را به سمت منفی بودن سوق دهد. به‌علاوه، تعداد کلمات در جملات منفی بیشتر است؛ پس احتمال منفی بودن احساسات بالا است و غیره.

1. Ontology Web Language

رابطه (۱) بیانگر مجموعه‌ای از قوانین تولیدشده آنتولوژیکی است که j تعداد قوانین تولیدشده را نشان می‌دهد.

$$O^R = \{O_1^R . O_2^R . \dots . O_j^R\}; 1 \le i \le j \qquad (۱)$$

قوانین به روش‌های مختلفی برقرار می‌شوند که تعدادی از آن‌ها در روابط (۲) نشان داده شده است

$$\begin{aligned} O_1^R &= \{C \perp S^C \perp T\} \\ O_2^R &= \{C \perp T\} \\ O_3^R &= \{S^C \perp T\} \\ O_4^R &= \{T_1 \perp T_2\} \end{aligned} \qquad (۲)$$

که درآن، $\perp$ نشان‌دهنده رابطه در آنتولوژی و T_1 و T_2 هر دو نمونه را در مجموعه T_i نشان می‌دهند. هر ویژگی $w_j \in O_j^R$، با معیاری که شباهت ویژگی j ام w_j را با قانون O_j^R ارزیابی می‌کند، مورد محاسبه قرار می‌گیرد. ویژگی مورد نیاز براساس تعداد ویژگی‌های V مطابق با رابطه (۳) انتخاب می‌شود.

$$m(w_j) = (\frac{T^j}{V}) \qquad (۳)$$

به‌طوری‌که $T^j = \sum_{j=1}^{i} T_i^j$، که در آن تابع T_i^j مقدار ۱ را در نظر می‌گیرد، اگر قانون آن متعلق به j امین ویژگی w_j باشد که

$$T_i^j = \begin{cases} 1\,; & if\ r_i\ belongs\ to\ w_j \\ 0\,; & Otherwise \end{cases} \qquad (۴)$$

زمانی‌که که بهترین ویژگی‌ها براساس امتیاز به‌دست آمده توسط رابطه (۲) انتخاب می‌شوند، بعد پایگاه داده YP به $n \times r$ از $n \times m$ کاهش می‌یابد که ورودی اعمال‌شده به یک الگوریتم آتوماتای یادگیر عمیق سلولی جهت تشخیص احساسات وارد می‌شود.

۳-۳- تحلیل احساسات با استفاده از آتوماتای یادگیر عمیق سلولی

این فاز از مدل پیشنهادی دارای دو بخش کلی می‌باشد که عبارتند از آتوماتای یادگیر سلولی و شبکه عصبی عمیق GMDH که در ادامه به تشریح هر کدام پرداخته می‌شود.

۳-۳-۱- آتوماتای یادگیری سلولی

اتوماتای یادگیر سلولی مدلی برای سیستم‌هایی است که از اجزاء ساده‌ای تشکیل شده‌اند و رفتار هر جزء براساس رفتار همسایگانش و نیز تجربیات گذشته‌اش تعیین و اصلاح می‌شود. اجزاء ساده تشکیل‌دهنده این مدل، از

طریق کنش و واکنش با یکدیگر می‌توانند رفتار پیچیده‌ای از خود نشان دهند. روال عملکرد اتوماتای یادگیر سلولی به‌این صورت است که مجموعه‌ای از جملات را به‌عنوان ورودی دریافت کرده و به بخش‌های کوچک‌تر تقسیم‌بندی می‌کند. پس از انطباق دامنه و تولید مدل‌های گرافی مبتنی‌بر اجزاء کلام، باید یک موتور پردازشی تولید شود که بتواند به پردازش این مدل‌ها و استخراج احساس و یا نظر نهفته در متن بپردازد. در این مقاله، برای پیدا کردن تطابق‌های الگویی کارا به‌منظور تشخیص احساس از یک مکانیزم ابداعی ترکیبی به‌نام آتوماتای یادگیر عمیق سلولی مبتنی‌بر شبکه عصبی عمیق GMDH یا بهره گرفته شده است.

در واقع D-CLA یک اتوماتای یادگیر سلولی یا CLA است که یادگیری آن مبتنی‌بر مفهوم یادگیری عمیق می‌باشد. از سوی دیگر CLA، مجموعه‌ای از سلول‌هاست که هر سلول آن، خودش یک اتوماتای یادگیر می‌باشد. مهم‌ترین خاصیت CLA یادگیری آن در تعامل با محیط می‌باشد. در CLA، یک مسأله بزرگ به یک سری مسأله‌های جزئی تبدیل می‌شود که هر جزء آن یک سلول بوده و هر سلول یک اتوماتای یادگیر مستقل است و می‌تواند برای یادگیری از اطلاعات سلول‌های همسایه خود نیز بهره ببرد. با توجه به اثبات کارایی یادگیری عمیق، در طرح پیشنهادی، یادگیری هر اتوماتای یادگیر مبتنی‌بر مفاهیم یادگیری عمیق خواهد بود تا دقت یادگیری CLA افزایش یابد. سپس هر گره از مدل گرافی تولیدشده از مرحله قبل، به یک سلول نگاشت‌شده و در تکامل یادگیری سلول‌های در تعامل با یکدیگر، احساس و یا نظر نهفته در متن استخراج می‌شود [۲۵].

۲-۳-۳- شبکه عصبی عمیق GMDH

روش GMDH اولین بار توسط ایواخنکو [۲۶] به‌صورت یک روش آنالیز چندمتغیره برای مدل‌سازی و شناسایی سامانه‌های پیچیده توسعه پیدا کرد. هدف اصلی GMDH بنا نهادن یک تابع تحلیلی براساس یک شبکه پیش‌خور[1] است که هر عنصر آن یک تابع درجه دوم شکل می‌دهد که ضرایب آن به‌کمک روش‌های رگرسیونی به‌دست می‌آیند [۲۷].

مطابق الگوریتم GMDH یک مدل می‌تواند به‌صورت مجموعه‌ای از نرون‌ها به‌طوری‌که زوج‌های مختلف آن‌ها در هر لایه از طریق یک معادله چندجمله‌ای مرتبه دوم به یکدیگر مرتبط شده‌اند، بیان شود و نرون‌های جدید در لایه‌های بعدی را بسازند. این چنین نمایشی می‌تواند در مدل‌سازی مسیر ورودی‌ها به خروجی به‌کار گرفته شود. تعریف مشخص و قراردادی شناسایی یک مسأله، یافتن تابع $\hat{f}$ است، به‌طوری‌که می‌تواند تقریباً به‌جای مقدار واقعی f به‌منظور پیش‌بینی خروجی $\hat{y}$ به‌ازای یک بردار ورودی مشخص $X = (x_1.x_2.x_3.\dots.x_n)$ در نزدیک‌ترین حالت ممکن به مقدار واقعی خروجی y به‌کار رود. بنابراین، M مشاهده مشخص از زوج داده‌های چند ورودی-یک خروجی به طوری که مطابق رابطه (۵):

$$y_i = f(x_{i1}.x_{i2}.x_{i3}.\dots.x_{in}), (i = 1,2,\dots,M) \quad (۵)$$

حال امکان آموزش یک شبکه عصبی از نوع GMDH برای پیش‌بینی مقادیر خروجی $\hat{y}_i$ برای هر بردار ورودی مشخص $X = f(x_{i1}.x_{i2}.x_{i3}.\dots.x_{in})$ وجود دارد. به عبارتی مطابق رابطه (۶) خواهیم داشت:

$$\hat{y}_i = \hat{f}(x_{i1}.x_{i2}.x_{i3}.\dots.x_{in}), (i = 1,2,\dots,M) \quad (٦)$$

اکنون مسأله تعیین یک شبکه عصبی GMDH به‌گونه‌ای است که مربع اختلافات بین مقدار واقعی خروجی و مقدار پیش‌بینی شده متناظر به حداقل برسد و به عبارتی مطابق رابطه (۷) داریم:

$$\sum_{i=1}^{M}[\hat{f}x_{i1}.x_{i2}.x_{i3}.\dots.x_{in}) - y_i]^2 \to min \quad (۷)$$

ارتباطات کلی بین ورودی‌ها و متغیرهای خروجی را می‌توان با استفاده از یک فرم پیچیده گسسته از سری توابع ولترا به فرم رابطه (۸) بیان نمود.

$$y = a_0 + \sum_{1}^{n} a_i x_i + \sum_{1}^{n}\sum_{1}^{n} a_{ij}x_i x_i + \sum_{1}^{n}\sum_{1}^{n}\sum_{1}^{n} a_{ijk}x_i x_j x_k + \cdots \quad (۸)$$

که به چندجمله‌ای Kolmogorov–Gabor شناخته می‌شوند. این فرم کامل از بیان ریاضی را می‌توان به‌صورت یک سامانه از چندجمله‌های مرتبه دو جزئی بیان نمود، به‌طوری‌که شامل تنها دو متغیر (نرون) در فرم رابطه (۹) باشند.

1 .Feed Forward Networks

$$\hat{y} = G(x_i.x_j) = a_0 + a_1x_i + a_2x_j + a_3x_ix_j + a_4x_j^2 + a_5x_j^2 \quad (۹)$$

که ضریب a_i در معادله (۹) به کمک روش‌های رگرسیونی محاسبه می‌شوند، به‌طوری‌که اختلاف بین مقدار واقعی (y) و مقدار برآورده‌شده ($\hat{y}$) برای هر زوج از متغیرهای ورودی X_i . X_j حداقل شود. در این راستا، ضرایب هر معادله مرتبه دو G_i برای برازش بهینه خروجی درکل مجموعه زوج داده‌های ورودی-خروجی به‌دست می‌آید. به عبارت دیگر مطابق رابطه (۱۰) خواهیم داشت.

$$E = \frac{\sum_{i=1}^{M}(y_i - G_i)^2}{M} \to min \quad (۱۰)$$

در فرم اصلی الگوریتم GMDH، کلیه حالات ممکن دو متغیر مستقل از مجموع n متغیر ورودی به‌منظور تشکیل چندجمله‌ای‌های رگرسیون مطابق رابطه (۵) در نظر گرفته می‌شوند، به‌طوری‌که بهترین برازش‌های مشاهدات وابسته (y_i . i=1,2,...,M) در جهت ارضای کمترین مربعات حاصل گردد. در نتیجه $\binom{n}{2} = \frac{n(n-1)}{2}$ نرون در اولین لایه از شبکه پیش‌خور از مشاهدات $\{(y_i.x_{ip}.x_{iq}), (i = 1,2,\dots,M)\}$ برای مقادیر مختلف $p.q \in \{i = 1,2,\dots,n\}$ گسترش می‌یابند. به‌عبارت دیگر، در این حالت امکان تشکیل M داده سه‌گانه $\{(y_i.x_{ip}.x_{iq}), (i = 1,2,\dots,M)\}$ از مشاهدات را مطابق رابطه (۱۱) به کمک چنین فرمی $p.q \in \{i = 1,2,\dots,n\}$ وجود دارد:

$$\begin{bmatrix} x_{1p} & x_{1q} & \vdots & y_1 \\ x_{2p} & x_{2q} & \vdots & y_2 \\ \cdots & \cdots & \vdots & \cdots \\ x_{Mp} & x_{Mq} & \vdots & y_M \end{bmatrix} \quad (۱۱)$$

به کمک زیرمجموعه مرتبه دو به فرم رابطه (۱۲) برای هر ردیف از M داده سه‌گانه، معادلات ماتریسی به آسانی به‌صورت زیر به دست می‌آید:

$$A_a = Y \quad (۱۲)$$

که در آن a بردار ضرایب ناشناخته از چندجمله‌ای‌های مرتبه دو معادله (۱۳) و (۱۴) است.

$$a = \{a_0.a_1.a_2.a_3.a_4.a_5\} \quad (۱۳)$$

$$Y = \{y_1.y_2.y_3.\dots.y_M\}^T \quad (۱۴)$$

بردار مقادیر خروجی مشاهدات است. مطابق رابطه (۱۵) می‌توان به آسانی دریافت که:

$$A = \begin{bmatrix} 1 & x_{1p} & x_{1q} & x_{1p}x_{1q} & x_{1p}^2 & x_{1q}^2 \\ 1 & x_{2p} & x_{2q} & x_{2p}x_{2q} & x_{2p}^2 & x_{2q}^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{Mp} & x_{Mq} & x_{Mp}x_{Mq} & x_{Mp}^2 & x_{Mq}^2 \end{bmatrix} \quad (۱۵)$$

روش حداقل مربعات از آنالیز رگرسیون چندمتغیر منجر به حل معادلات ضرایب به فرم رابطه (۱۶) می‌شود.

$$a = (A^TA)^{-1}A^TY \quad (۱۶)$$

این معادله بردار بهترین ضرایب برای کل مجموعه M داده سه‌گانه تعیین می‌کند. این روش برای هر نرون از لایه پنهان بعدی مطابق با توپولوژی شبکه تکرار می‌شود [28-30]. بنابراین، براساس روابط تعریف‌شده برای الگوریتم GMDH، فرآیند آموزش و تولید مدل جهت تحلیل و تشخیص احساسات کاربران در شبکه‌های اجتماعی صورت می‌گیرد.

۴- ارزیابی نتایج آزمایش

در این بخش، ابتدا مشخصات سخت‌افزاری و نرم‌افزاری شبیه‌سازی آزمایش بیان شده است. سپس، مجموعه داده‌های مورد آزمایش، معیارهای ارزیابی آزمایش و درنهایت، تحلیل نتایج آزمایش‌ها بررسی شده است.

۴-۱- مشخصات سخت‌افزاری و نرم‌افزاری شبیه‌سازی آزمایش

روش پیشنهادی و روش‌های مقایسه‌شده در این مقاله، با استفاده از شبیه‌ساز MATLAB-R2020b پیاده‌سازی شده است. سیستم عامل استفاده‌شده در محیط آزمایش، ویندوز ۷ نوع ۳۲ بیتی، حافظه RAM استفاده‌شده ۴ گیگابایت (۳,۰۶ گیگ قابل استفاده)، پردازنده اینتل Core™-i7 با ۷ هسته با مشخصات Q720-1.60GHz می‌باشد. لازم به ذکر است که نتایج بیان شده در آزمایش‌ها، حاصل یک حلقه با ۲۰ مرتبه اجرای الگوریتم و سپس میانگین‌گیری نتایج نشان داده شده است. بنابراین با توجه به یک محیط خاص سخت‌افزار و نرم‌افزاری با مشخصات فوق، شبیه‌سازی این تحقیق انجام‌شده و نتایج ارزیابی شده است.

۲-۴- مجموعه داده‌های مورد آزمایش

مجموعه داده‌های مورد آزمایش در این مقاله، به دو صورت دسته‌ای[1] و جریانی[2] است. در این راستا، استفاده از مجموعه‌داده‌های برخط موجود برای بارگیری داده از شبکه‌های اجتماعی از راه‌کارهای اساسی برای مجموعه داده‌های جریانی می‌باشد. همچنین با توجه به این‌که احساس، عقیده و نظر در هر زبان و گویشی وجود دارد، یک مدل کارآمد باید بتواند صرف‌نظر از زبان، به استخراج نظر، عقیده و احساس موجود در آن بپردازد. به‌علاوه، برای بررسی مجموعه داده‌های دسته‌ای، از پایگاه‌داده مشتریان آمازون در بخش اول آزمایش‌ها استفاده شده است [۱۷]. به‌طورکلی، در این مقاله از مجموعه داده‌های مشترک در مقاله [۱۷] استفاده شده و با نتایج به‌دست آمده مورد مقایسه قرار گرفته است.

در بخش دوم آزمایش‌ها، جهت اثبات درستی و عملکرد روش پیشنهادی، به بررسی عملکرد آن بر روی مجموعه داده‌های دیگر پرداخته می‌شود. در این مقاله، از پنج مجموعه داده معروف و رایج استفاده شده است که عبارتند از توئیتر [۱۸]، فیس‌بوک [۱۹]، اخبار جعلی COVID-19 [20]، آمازون [21] و شبکه اخبار جعلی [۲۲]. توئیتر یک مجموعه داده sentiment140 و شامل ۱۶۰۰۰۰۰ توئیت است که با استفاده از API توئیتر استخراج شده است. توئیت‌ها توسط مقادیر ۰ (منفی) الی ۴ (مثبت) حاشیه‌نویسی شده‌اند که می‌توان ازآن‌ها برای تشخیص احساسات استفاده کرد. به‌علاوه، مجموعه داده توئیتر شامل شش ویژگی است که در جدول (۲) نشان داده شده است [۱۸].

جدول ۲. مشخصات ویژگی‌های مجموعه‌داده توئیتر

ویژگی	توصیف ویژگی
هدف	قطبیت توئیت (۰ : منفی، ۲ : خنثی، ۴ : مثبت)
شناسه	شناسه توئیت (۲۰۸۷)
تاریخ	تاریخ توئیت (شنبه ۱۶ مه ۲۳:۵۸:۴۴ UTC 2009)
پرچم	پرس‌وجو (lyx). اگر درخواستی وجود نداشته باشد، این مقدار NO_QUERY است.
کاربر	کاربری که توئیت کرده است (robotickilldozr)
متن	متن توئیت (Lyx جالب است)

مجموعه داده فیس‌بوک [۱۹] شامل ۱۰۰۰۰ مقاله خبری و متاداده تقریباً از بین‌رفته است. ۶۰۰ صفحه وب از وب‌سایت Politifact برای تجزیه و تحلیل آن با استفاده از مهارت‌های علم داده و دریافت برخی بینش‌ها در مورد این‌که چگونه می‌توان از انتشار اطلاعات نادرست در جنبه گسترده‌تری جلوگیری نمود و چه رویکردی دقت بهتری برای رسیدن به آن می‌دهد. این مجموعه داده دارای شش ویژگی است که در جدول (۳) نشان داده شده است و در میان آن‌ها، ویژگی عنوان اخبار برای طبقه‌بندی اخبار به‌عنوان غلط یا درست مهم‌ترین آن‌ها می‌باشد.

1. Batch
2 .Stream

جدول ۳. مشخصات ویژگی‌های مجموعه‌داده فیس‌بوک

ویژگی	توصیف ویژگی
عنوان اخبار	حاوی اطلاعاتی که باید مورد تجزیه و تحلیل قرار گیرد.
پیوند اخبار	حاوی URL سرفصل‌های اخبار مشخص شده در ویژگی اول.
منبع	نام نویسندگانی که اطلاعات را در فیس‌بوک، اینستاگرام، توئیتر یا هر پلتفرم رسانه اجتماعی دیگری ارسال کرده‌اند.
تاریخ اعلام	حاوی تاریخ ارسال اطلاعات توسط نویسندگان در پلتفرم‌های مختلف رسانه‌های اجتماعی.
تاریخ بررسی	حاوی تاریخی است که این بخش از اطلاعات توسط تیم بررسی‌کننده واقعیت Polyfact تجزیه و تحلیل شده تا به‌عنوان جعلی یا واقعی برچسب‌گذاری شود.
برچسب	پنج برچسب کلاس شامل درست، اکثراً درست، نیمه‌درست، به‌سختی درست، غلط و غیره.

مجموعه داده اخبار جعلی COVID-19 [۲۰]، نتیجه یک وظیفه فرعی در کار مشترک CONSTRAINT-2021 در تشخیص پست متخاصم است. این کار فرعی بر شناسایی اخبار جعلی مرتبط با COVID-19 به زبان انگلیسی متمرکز است. منابع داده‌ها، پلتفرم‌های رسانه‌های اجتماعی مختلف مانند توئیتر، فیس‌بوک، اینستاگرام و غیره هستند. با توجه به یک پست رسانه‌های اجتماعی، هدف از کار مشترک طبقه‌بندی آن به دو دسته اخبار جعلی یا واقعی است.

مجموعه داده آمازون [۲۱]، برای مقاله "از گروه به برچسب‌های فردی با استفاده از ویژگی‌های عمیق" به نویسندگی Kotzias et. al، در سال ۲۰۱۵ ایجاد شده است و حاوی جملاتی است که دارای برچسب احساسات مثبت یا منفی هستند. امتیاز یا ۱ برای مثبت و یا ۰ برای منفی است. جملات از سه وب‌سایت شامل imdb.com، amazon.com و yelp.com در زمینه‌های مختلف جمع‌آوری شده است. برای هر وب‌سایت ۵۰۰ جمله مثبت و ۵۰۰ جمله منفی وجود دارد که به‌طور تصادفی برای مجموعه داده‌های بزرگ‌تر بررسی‌ها انتخاب شدند. در این مقاله جملاتی انتخاب شده است که به وضوح مفهوم مثبت یا منفی داشته باشند، هدف این است که هیچ جمله خنثی انتخاب نشود.

وب‌سایت IMDB به مجموعه‌داده‌های احساسات نقد فیلم اشاره دارد که در ابتدا توسط Maas و همکارانش معرفی شد. به‌عنوان معیاری برای تحلیل احساسات، این مجموعه داده شامل ۱۰۰۰۰۰ نقد فیلم است که ۵۰۰۰۰ بررسی بدون برچسب وجود داشته و ۵۰۰۰۰ باقی‌مانده به مجموعه‌ای از ۲۵۰۰۰ بررسی برای آموزش و ۲۵۰۰۰ بررسی برای آزمایش تقسیم می‌شوند. هر یک از بررسی‌های برچسب‌گذاری‌شده دارای یک برچسب احساسات باینری است، مثبت یا منفی. در آزمایش‌های این مقاله، فقط روی قسمت برچسب‌گذاری شده مجموعه آموزشی انجام شده است. وب‌سایت آمازون حاوی نظرات و امتیازات محصولات فروخته شده در دسته تلفن‌های همراه و لوازم جانبی است و بخشی از مجموعه داده جمع‌آوری شده توسط McAuley و Leskovec می‌باشد. نمرات در مقیاس اعداد صحیح از ۱ تا ۵ است. در این مقاله، نظرات با نمره ۴ و ۵ مثبت و امتیازات ۱ و ۲ منفی در نظر گرفته شده‌اند. داده‌ها به‌طور تصادفی به دو نیمه ۵۰٪ تقسیم شده، یکی برای آموزش و دیگری برای آزمایش، با ۳۵۰۰۰ سند در هر مجموعه. وب‌سایت Yelp شامل مجموعه‌داده‌های نظرات در مورد رستوران‌هاست که در این مقاله استخراج شده است. نمرات در یک مقیاس اعداد صحیح از ۱ تا ۵ است. به‌طور مشابه، نظرات با نمرات ۴ و ۵ مثبت و ۱ و ۲ منفی در نظر گرفته شده‌اند. به‌طور تصادفی یک تقسیم آموزشی و آزمایشی ۵۰-۵۰ ایجاد که منجر به تقریباً ۳۰۰۰۰۰ سند برای هر مجموعه شد.

مجموعه داده شبکه اخبار جعلی [۲۲]، یک مخزن برای یک پروژه جمع‌آوری داده‌های جاری برای تحقیقات اخبار جعلی در ASU است و شامل کلیه مقالات اخبار جعلی با ویژگی‌های محتوای اخباری است که در جدول (۴) نشان داده شده است. برای هر یک از مجموعه داده‌های مذکور، ۱۰۰۰ جمله از مجموعه آزمایش استخراج‌شده و به‌صورت دستی بر چسب‌گذاری شده است که ۵۰٪ احساسات مثبت و ۵۰٪ احساسات منفی بودند. این جملات فقط برای ارزیابی طبقه‌بندی‌کننده سطح نمونه این مقاله برای هر مجموعه داده، مقدار ۳ استفاده می‌شوند. آن‌ها برای آموزش مدل، برای حفظ سازگاری با هدف کلی ما از یادگیری در سطح گروهی و پیش‌بینی در سطح نمونه استفاده نمی‌شوند [۱۷].

جدول ۴. مشخصات ویژگی‌های مجموعه‌داده شبکه اخبار جعلی

ویژگی	توصیف ویژگی
منبع	نویسنده یا ناشر خبر.
تیتر	متن کوتاهی که هدف آن جلب‌توجه خوانندگان بوده و با عمده موضوع خبری کاملا مرتبط است.
متن	جزئیات خبر را تشریح کرده و اغلب یک ادعای عمده وجود دارد که زاویه ناشر را شکل داده و به‌طور خاص برجسته و توضیح داده می‌شود.
تصویر-ویدئو	بخش مهمی از محتوای متن خبر است که نشانه‌های بصری را برای قاب‌بندی داستان ارائه می‌دهد.

۳-۴- معیارهای ارزیابی آزمایش

برای ارزیابی عملکرد سیستم‌های داده‌کاوی، معیارهای ارزیابی متفاوتی به‌کار گرفته می‌شوند. برای تشخیص موضع از معیارهای درستی (Acc[1])، دقت (Pre[2])، حساسیت (Rec[3]) و F_{score}، استفاده می‌شود که تعریف ریاضی هریک به‌ترتیب در روابط (۱۷) الی (۲۰) نشان داده شده است [۲۳]. در این روابط، a تعداد پست‌هایی است که از بحث پشتیبانی می‌کنند و به نفع آن‌ها پیش‌بینی می‌شود و c تعداد پست‌هایی است که مخالف بحث هستند، اما موافق بحث در نظر گرفته می‌شوند. به‌طور مشابه، b تعداد پست‌هایی است که مخالف این بحث هستند و مخالف بحث نیز در نظر گرفته می‌شوند و d تعداد پست‌هایی است که موافق بحث هستند، اما مخالف آن برچسب‌گذاری می‌شوند [۲۴].

$$Accuracy = \frac{a+b}{a+b+c+d} \quad (۱۷)$$

$$Precision = \frac{a}{a+c} \quad (۱۸)$$

$$Recall = \frac{a}{a+d} \quad (۱۹)$$

به‌علاوه، یکی از مهم‌ترین معیارهای ارزیابی در این مقاله مرتبه زمانی می‌باشد. مرتبه زمانی براساس میلی‌ثانیه و با استفاده از رابطه (۲۰) محاسبه می‌گردد.

$$Execuation\ Time = \sum_{i=1}^{i} t \quad (۲۰)$$

به‌علاوه، یکی از مهم‌ترین معیارهای ارزیابی در این مقاله مرتبه زمانی می‌باشد. مرتبه زمانی براساس میلی‌ثانیه و با استفاده از رابطه (۲۰) محاسبه می‌گردد.

$$Execuation\ Time = \sum_{i=1}^{i} t \quad (۲۰)$$

۴-۴- تحلیل و ارزیابی نتایج آزمایش‌ها

در جدول (۵)، مقایسه روش پیشنهادی بر روی مجموعه داده آمازون براساس معیارهای دقت، صحت، حساسیت و خطای تحلیل احساسات بدون اعمال و با اعمال آنتولوژی نشان داده شده است.

1. Accuracy
2.Precision
3. Recall

جدول ۵. مقایسه روش پیشنهادی بر روی مجموعه داده آمازون براساس معیارهای دقت، صحت، حساسیت و خطای تحلیل احساسات بدون اعمال و با اعمال آنتولوژی

روش پیشنهادی با آنتولوژی				CSO-LSTMNN [36]			
Accuracy	Recall	Precision	Error	Accuracy	Recall	Precision	Error
٪**98.2**	٪**73.83**	٪**74.21**	٪**1.8**	٪96.89	٪72	٪73	٪3.11

همان‌طور که مشاهده می‌گردد، میزان بهبود دقت روش پیشنهادی مبتنی‌بر آنتولوژی نسبت‌به روش پیشنهادی مبتنی‌بر شبکه عصبی عمیق سلولی بدون اعمال آنتولوژی برابر با ۴,۷٪ می‌باشد. میزان بهبود حساسیت برابر با ۱٪، میزان بهبود صحت روش پیشنهادی معنایی نسبت‌به میزان صحت روش پیشنهادی بدون اعمال آنتولوژی برابر با ۳٪ و. میزان بهبود خطای روش پیشنهادی معنایی نسبت‌به میزان خطای روش پیشنهادی بدون اعمال آنتولوژی برابر با ۴٪ می‌باشد.

از نتایج به‌دست آمده در جدول (۵) می‌توان به این نتیجه رسید که روش پیشنهادی مبتنی‌بر معنایی در مقایسه با روش پیشنهادی بدون استفاده از آنتولوژی بهتر عمل می‌کند. دلیل بهبود روش معنایی نسبت‌به روش غیره معنایی این است که آنتولوژی، براساس تحلیل مفاهیم معنایی داده‌ها، مدل ساده‌تری را تولید می‌کند. درنهایت با وجود یک مدل ساده، داده‌های آزمایش به دقت بالاتری طبقه‌بندی می‌شوند. در روش پیشنهادی هرچه حجم نمونه‌ها جهت آموزش شبکه عصبی عمیق سلولی بیشتر باشد، مدل دقیق‌تری تولید می‌گردد. با توجه به نتایج به‌دست آمده، مشاهده می‌گردد که استفاده از روش معنایی به میزان قابل قبولی می‌تواند دقت تحلیل احساسات نسبت‌به حالت معمولی و سایر روش‌ها را براساس تعداد داده‌های متنوع بهبود بخشد.

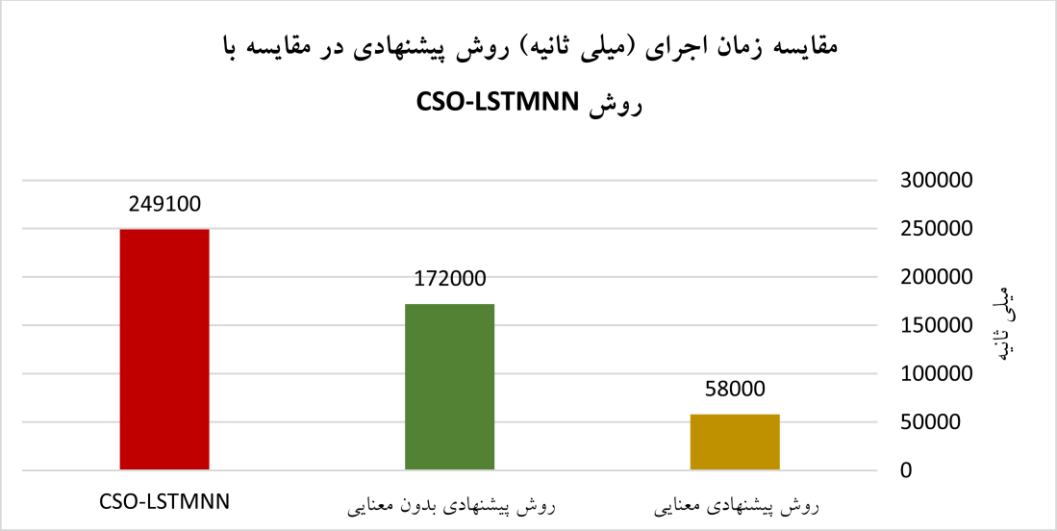
در جدول (۶)، مقایسه عملکرد روش پیشنهادی بر روی مجموعه داده آمازون براساس معیارهای دقت، صحت، حساسیت و خطا نسبت‌به سایر روش‌های دیگر نشان داده شده است.

جدول ۶. مقایسه عملکرد روش پیشنهادی براساس معیارهای دقت، صحت، حساسیت و خطا نسبت به سایر روش CSO-LSTMNN

روش پیشنهادی با استفاده از آنتولوژی				روش پیشنهادی بدون آنتولوژی			
Accuracy	Recall	Precision	Error	Accuracy	Recall	Precision	Error
٪98.2	٪73.83	٪74.83	٪1.8	٪94.72	٪72.83	٪70.83	٪5.28

همان‌طور که مشاهده می‌گردد، میزان بهبود دقت روش پیشنهادی نسبت‌به میزان دقت، فراخوانی، حساسیت و خطای روش CSO-LSTMNN به‌ترتیب برابر با ۱,۳۱٪، ۱,۸۳٪، ۱,۲۱٪ و ۱,۳۱٪ می‌باشد.

با توجه به نتایج مقایسه‌ای در جدول (۶) می‌توان به این نتیجه رسید که رویکرد معنایی در فرآیند نظرکاوی با کمک آتوماتای یادگیر عمیق سلولی توانسته است به میزان قابل توجه‌ای بر روی داده‌های بزرگ نسبت‌به روش[۳۶] CSO-LSTMNN عمل کند. هرچه میزان و حجم نمونه‌ها افزایش یابد، خطای نظرکاوی کاهش یافته و به مراتب به یک مقدار قابل قبول و مطلوب می‌رسد. در شکل (۳)، مقایسه زمان اجرای روش پیشنهادی در مقایسه با سایر روش‌های دیگر بر روی مجموعه داده آمازون نشان داده شده است



شکل ۳. مقایسه زمان اجرای (میلی ثانیه) روش پیشنهادی در مقایسه با روش CSO-LSTMNN

با توجه به نتایج به‌دست آمده از شکل (۳) مشاهده می‌گردد که در کلیه روش‌ها، با افزایش حجم داده‌ها، زمان اجرا افزایش می‌یابد؛ اما برترین عملکرد زمانی (۶۱۰۰۰ میلی‌ثانیه) را روش پیشنهادی با اعمال رویکرد آنتولوژی دارد. علت برتری زمانی آنتولوژی این است که داده‌ها به مقادیر معنادار تبدیل شده و مدل ساده‌تری را ایجاد می‌نماید.

همان‌طور که بیان شد، در بخش دوم آزمایش‌ها جهت اثبات درستی و عملکرد روش پیشنهادی، به بررسی عملکرد آن بر روی پنج مجموعه داده معروف و رایج شامل توئیتر، فیس‌بوک، اخبار جعلی COVID-19، آمازون و شبکه اخبار جعلی پرداخته شده است.

جدول (۷)، نتایج معیار دقت مدل پیشنهادی را با روش‌های دیگر ازجمله شبکه عصبی عمیق حافظه کوتاه مدت (LSTM)، شبکه عصبی عمیق عودکننده (RDNN[1])، شبکه عصبی پرسپترون (MLP[2])، درخت تصمیم C4.5، نزدیک‌ترین همسایه (KNN[3]) و ماشین بردار پشتیبان (SVM) مقایسه کرده است.

جدول ۷. مقایسه نتایج دقت مدل پیشنهادی با روش‌های LSTM، RDNN، MLP، C4.5، KNN و SVM

	توئیتر ٪	فیس‌بوک ٪	COVID-19 ٪	آمازون ٪	شبکه اخبار جعلی ٪
Proposed			**90**	**98.20**	**97.20**
			80	96.00	95.00
			95	95.12	94.87
			57	92.34	91.41
			16	92.64	91.03
			81	90.94	92.69
			61	89.94	88.44

1. Recurrent Neural Network
2 .Multi Layer Perceptron
3. K-Nearest Neighbor

همان‌طور که از جدول (۷) مشاهده می‌شود، میزان بهبود دقت نظرکاوی روی مجموعه داده توئیتر در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ۲٪، ۲,۵٪، ۴,۹۲٪، ۵,۰۱٪، ۶,۹۱٪ و ۷,۷۴٪ می‌باشد. همچنین، میزان بهبود دقت نظرکاوی روی مجموعه داده فیس‌بوک در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ۲,۴٪، ۳,۰۶٪، ۶,۴۴٪، ۶,۸۲٪، ۵,۱۱٪ و ۹,۳۹٪ است. به‌علاوه، میزان بهبود دقت نظرکاوی روی مجموعه داده اخبار جعلی COVID-19 در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ۲,۱٪، ۲,۹۵٪، ۶,۳۳٪، ۶,۷۴٪، ۵,۰۹٪ و ۹,۲۹٪ می‌باشد. به‌طور مشابه، میزان بهبود دقت نظرکاوی روی مجموعه داده آمازون در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ۲,۲٪، ۳,۰۸٪، ۵,۸۶٪، ۵,۵۶٪، ۵,۸۶٪ و ۸,۲۶٪ است. درنهایت، میزان بهبود دقت نظرکاوی روی مجموعه داده شبکه اخبار جعلی در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ۲,۲٪، ۲,۳۳٪، ۵,۷۹٪، ۶,۱۷٪، ۴,۵۱٪ و ۸,۷۶٪ می‌باشد.

با توجه به نتایج به‌دست آمده در جدول (۷) مشاهده گردید که روش پیشنهادی دارای عملکرد بهتری است، چراکه اعمال رویکرد معنایی با یک سطح بهینه و همچنین، مدل شبکه عصبی عمیق GMDH سلولی به میزان قابل توجه‌ای توانسته است بهتر عمل کرده و صحت قابل قبولی نسبت‌به سایر روش‌های دیگر داشته باشد.

جدول (۸)، نتایج معیار صحت مدل پیشنهادی را با نسبت‌به روش‌های دیگر از جمله LSTM، RDNN، MLP، C4.5، KNN و SVM مقایسه کرده است.

جدول ۸. مقایسه نتایج صحت مدل پیشنهادی با روش‌های LSTM، RDNN، MLP، C4.5، KNN و SVM

	توئیتر ٪	فیس‌بوک ٪	COVID-19 ٪	آمازون ٪	شبکه اخبار جعلی ٪
Proposed Model+Ontology	**99.15**	**98.65**	**98.60**	**99.60**	**98.80**
LSTM	98.96	98.08	98.08	98.54	98.06
RDNN	98.43	97.51	97.50	97.80	97.46
MLP	95.79	93.82	93.80	94.76	93.68
C4.5	95.06	92.81	92.77	94.44	92.65
KNN	92.86	94.44	94.41	92.44	94.32
SVM	91.59	89.34	89.31	90.96	89.13

همان‌طور که در جدول (۸) مشاهده می‌شود، میزان بهبود صحت نظرکاوی روی مجموعه داده توئیتر در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ۰,۱۹٪، ۰,۷۲٪، ۳,۳۶٪، ۴,۰۹٪، ۴,۲۹٪ و ۷,۵۶٪ می‌باشد. همچنین، میزان بهبود صحت نظرکاوی روی مجموعه داده فیس‌بوک در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ۰,۵۷٪، ۱,۱۴٪، ۴,۸۳٪، ۵,۸۴٪، ۴,۲۱٪ و ۹,۳۱٪ است. به‌علاوه، میزان بهبود صحت نظرکاوی روی مجموعه داده اخبار جعلی COVID-19 در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ۰,۵۲٪، ۱,۱٪، ۴,۸٪، ۵,۸۳٪، ۴,۱۹٪ و ۹,۲۹٪ می‌باشد. به‌طور مشابه، میزان بهبود صحت نظرکاوی روی مجموعه داده آمازون در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ۱,۰۶٪، ۱,۸٪، ۴,۸۴٪، ۵,۱۶٪، ۷,۱۶٪ و ۸,۶۴٪ است. درنهایت، میزان بهبود صحت نظرکاوی روی مجموعه داده شبکه اخبار جعلی در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM

به‌ترتیب برابر با ۰. ۷۴٪، ۱,۳۴٪، ۵,۱۲٪، ۶,۱۵٪، ۴,۴۸٪ و ۹,۶۷٪ می‌باشد. مشخص شد که مدل پیشنهادی بهتر از سایر روش‌ها عمل کرده است.

با توجه به نتایج به‌دست آمده از مدل پیشنهادی در جدول (۸) مشاهده گردید که مدل تولیدشده مبتنی‌بر آنتولوژی، ساده و با دقت بالا عمل می‌کند. ازاین‌رو، ساده بودن مدل با تعداد لایه‌های پنهان بالا موجب شده است که صحت نظرکاوی افزایش یابد.

جدول (۹)، نتایج معیار حساسیت مدل پیشنهادی را با روش‌های دیگر از جمله LSTM، RDNN، MLP، C4.5، KNN و SVM مقایسه کرده است. همان‌طور که مشاهده می‌شود، میزان بهبود حساسیت نظرکاوی روی مجموعه داده توئیتر در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ٪۲,۳۶، ٪۲,۳۹، ٪۲,۳۸، ٪۱,۷۷، ٪۱,۶ و ٪۱,۱۹ می‌باشد. همچنین، میزان بهبود حساسیت نظرکاوی روی مجموعه داده فیس‌بوک در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ٪۲,۳۵، ٪۲,۴۹، ٪۲,۵۸، ٪۱,۹۷، ٪۱,۷۹ و ٪۱,۳۸ است. به‌علاوه، میزان بهبود حساسیت نظرکاوی روی مجموعه داده اخبار جعلی COVID-19 در روش پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ٪۲,۱۵، ٪۲,۵، ٪۲,۵۸، ٪۱,۹۸، ٪۱,۸ و ٪۱,۳۸ می‌باشد. به‌طور مشابه، میزان بهبود حساسیت نظرکاوی روی مجموعه داده آمازون در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ٪۱,۳، ٪۱,۷۱، ٪۱,۸۸، ٪۲,۵، ٪۲,۵۱ و ٪۲,۲۸ است. درنهایت، میزان بهبود حساسیت نظرکاوی روی مجموعه داده شبکه اخبار جعلی در روش پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5 و KNN به‌ترتیب برابر با ٪۰,۷۹، ٪۱,۵۹، ٪۱,۱۴، ٪۱,۱۳، ٪۰,۶۱ و ٪۰,۴۳ می‌باشد.

جدول ۹.مقایسه نتایج حساسیت مدل پیشنهادی با روش‌های LSTM، RDNN، MLP، C4.5، KNN و SVM

	توئیتر ٪	فیس‌بوک ٪	COVID-19 ٪	آمازون ٪	شبکه اخبار جعلی ٪
Proposed Model+Ontology	**99.50**	**99.60**	**99.60**	**99.60**	97.40
LSTM	97.14	97.25	97.45	97.32	96.60
RDNN	97.11	97.11	97.10	97.09	97.05
MLP	97.12	97.02	97.02	97.10	96.96
C4.5	97.73	97.63	97.62	97.72	97.58
KNN	97.90	97.81	97.80	97.89	97.76
SVM	98.31	98.22	98.22	98.30	**98.19**

با توجه به نتایج به‌دست آمده، مشاهده گردید که روش پیشنهادی دارای عملکرد بهتری است؛ زیرا اعمال رویکرد معنایی با یک سطح بهینه و همچنین مدل شبکه عصبی عمیق GMDH سلولی به میزان قابل توجه‌ای توانسته است بهتر عمل کرده و صحت قابل قبولی نسبت‌به سایر روش‌ها داشته باشد.

جدول (۱۰)، نتایج معیار خطای مدل پیشنهادی را با روش‌های دیگر از جمله LSTM، RDNN، MLP، C4.5، KNN و SVM مقایسه کرده است. همان‌طور که مشاهده می‌شود، میزان بهبود خطای نظرکاوی روی مجموعه داده توئیتر در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ٪۲,۹، ٪۳,۴، ٪۵,۸۲، ٪۵,۹۱، ٪۷,۸۱ و ٪۸,۶۴ می‌باشد. همچنین، میزان بهبود خطای نظرکاوی روی مجموعه داده فیس-بوک در مدل پیشنهادی، در مقایسه با سایر روش‌های

دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ٪۳,۴، ٪۴,۰۶، ٪۷,۴۴، ٪۷,۸۲، ٪۶,۱۱ و ٪۱۰,۳۹ است. به‌علاوه، میزان بهبود خطای نظرکاوی روی مجموعه داده اخبار جعلی COVID-19 در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ٪۳,۲، ٪۴,۰۵، ٪۷,۴۳، ٪۷,۸۴، ٪۶,۱۹ و ٪۱۰,۳۹ می‌باشد. به‌طور مشابه، میزان بهبود خطای نظرکاوی روی مجموعه داده آمازون در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ٪۳,۲، ٪۴,۰۸، ٪۶,۸۶، ٪۶,۵۶، ٪۸,۲۶ و ٪۹,۲۶ است. درنهایت، میزان بهبود خطای نظرکاوی روی مجموعه داده شبکه اخبار جعلی در مدل پیشنهادی، در مقایسه با سایر روش‌های دیگر شامل LSTM، RDNN، MLP، C4.5، KNN و SVM به‌ترتیب برابر با ٪۲,۲، ٪۲,۳۳، ٪۵,۷۹، ٪۶,۱۷، ٪۴,۵۱ و ٪۸,۷۶ می‌باشد.

جدول ۱۰. مقایسه نتایج خطای مدل پیشنهادی با روش‌های LSTM، RDNN، MLP، C4.5، KNN و SVM

	توئیتر ٪	فیس‌بوک ٪	COVID-19 ٪	آمازون ٪	شبکه اخبار جعلی ٪
Proposed Model+Ontology	**1.80**	**2.00**	**2.10**	**1.80**	**2.80**
LSTM	3.80	4.40	4.20	4.00	5.00
RDNN	4.30	5.06	5.05	4.88	5.13
MLP	6.72	8.44	8.43	7.66	8.59
C4.5	6.81	8.82	8.84	7.36	8.97
KNN	8.71	7.11	7.19	9.06	7.31
SVM	9.54	11.39	11.39	10.06	11.56

با توجه به نتایج به‌دست آمده از مدل پیشنهادی در جدول (۱۰)، مشاهده گردید که مدل تولیدشده مبتنی‌بر آنتولوژی، ساده و با نرخ خطای پایین عمل می‌کند. ازاین رو، ساده بودن مدل با تعداد لایه‌های پنهان بالا موجب شده است که خطای نظرکاوی کاهش یابد.

۵- نتیجه‌گیری و کارهای آینده

از اهداف اصلی این مقاله، ارائه یک مدل نظرکاوی معنایی مبتنی‌بر آنتولوژی با استفاده از آتوماتای یادگیر عمیق سلولی مبتنی‌بر شبکه عصبی عمیق GMDH است. در این مقاله، از رویکرد معنایی آنتولوژی و آتوماتای یادگیر عمیق سلولی برای تحلیل و تشخیص نوع احساسات کاربران در انواع شبکه‌های اجتماعی استفاده شده است. از رویکرد آنتولوژی برای انتخاب ویژگی‌های برجسته براساس پایگاه دانش قوانین تولیدشده و از آتوماتای یادگیر عمیق سلولی مبتنی‌بر شبکه عصبی عمیق GMDH برای تحلیل و تشخیص احساسات کاربران در شبکه‌های اجتماعی استفاده شده است. با مشاهده نتایج به‌دست آمده، اثبات گردید که استفاده از آنتولوژی در فرآیند تشخیص احساسات و شناسایی توئیت‌های مثبت و منفی، طبقه‌بندی اخبار ویروس COVID-19، کامنت‌های کاربران و غیره موجب بهبود دقت شده است. استفاده از آتوماتای یادگیر عمیق سلولی مبتنی‌بر شبکه عصبی عمیق GMDH با قابلیت آموزش از توئیت‌های انجام شده، کامنت‌ها، اخبار و غیره موجب شد تا بتواند توئیت‌های جدید را براساس وزن‌دهی کلمات معنادار پیش‌بینی نموده و همین فرآیند موجب بهبود دقت تشخیص احساسات شده است. با شبیه‌سازی مدل پیشنهادی بر روی مجموعه داده‌های توئیتر، فیس‌بوک، اخبار جعلی COVID-19، آمازون و شبکه اخبار جعلی و ارائه رویکرد معنایی آنتولوژی در چارچوب اسکرام برای تشخیص احساسات، توانستیم به‌طور میانگین تا دقت 98.2٪ پیش رویم. ازاین‌رو، به‌کارگیری رویکرد آنتولوژی به میزان قابل توجه‌ای توانست سرعت اجرا و دقت تشخیص توسط رویکرد شبکه عصبی عمیق سلولی را بهبود بخشد.

استفاده از الگوریتم‌های شبکه عصبی عمیق همچون CNN، AutoEncoder و ترکیبی از روش‌های یادگیری ماشین با الگوریتم‌های بهینه‌سازی نظیر گرگ خاکستری، گربه‌های پیشرفته، سنجاقک و غیره به‌جای رویکرد آنتولوژی و شبکه عصبی عمیق سلولی برخی از مهم‌ترین پیشنهاداتی است که می‌توان جهت توسعه و بهبود نتایج این تحقیق ارائه نمود.

منابع

1.Yoo, S., Song, J., & Jeong, O. (2018). Social media contents-based sentiment analysis and prediction system. *Expert Systems with Applications*, *105*, 102-111.
2. Liu, B. (2020). Text sentiment analysis based on CBOW model and deep learning in big data environment. *Journal of ambient intelligence and humanized computing*, *11*(2), 451-458.
3. Singh, N. K., Tomar, D. S., & Sangaiah, A. K. (2020). Sentiment analysis: a review and comparative analysis over social media. *Journal of Ambient Intelligence and Humanized Computing*, *11*(1), 97-117.
4. Mandloi, L., & Patel, R. (2020, June). Twitter sentiments analysis using machine learninig methods. In *2020 International Conference for Emerging Technology (INCET)* (pp. 1-5). IEEE.
5. Chen, L. C., Lee, C. M., & Chen, M. Y. (2020). Exploration of social media for sentiment analysis using deep learning. *Soft Computing*, *24*(11), 8187-8197.
6. Chauhan, U. A., Afzal, M. T., Shahid, A., Abdar, M., Basiri, M. E., & Zhou, X. (2020). A comprehensive analysis of adverb types for mining user sentiments on amazon product reviews. World Wide Web, 23(3), 1811-1829.
7. Behera, R. K., Jena, M., Rath, S. K., & Misra, S. (2021). Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data. *Information Processing & Management*, *58*(1), 102435.
8. Chandra, S., Gourisaria, M. K., Harshvardhan, G. M., Rautaray, S. S., Pandey, M., & Mohanty, S. N. (2021). Semantic Analysis of Sentiments through Web-Mined Twitter Corpus. In ISIC (pp. 122-135).
9. Pathak, A. R., Pandey, M., & Rautaray, S. (2021). Topic-level sentiment analysis of social media data using deep learning. *Applied Soft Computing*, *108*, 107440.
10. Awajan, I., Mohamad, M., & Al-Quran, A. (2021). Sentiment analysis technique and neutrosophic set theory for mining and ranking big data from online reviews. *IEEE Access*, *9*, 47338-47353.
11. Cai, Y., Ke, W., Cui, E., & Yu, F. (2022). A deep recommendation model of cross-grained sentiments of user reviews and ratings. Information Processing & Management, 59(2), 102842.
12. Revathy, G., Alghamdi, S. A., Alahmari, S. M., Yonbawi, S. R., Kumar, A., & Haq, M. A. (2022). Sentiment analysis using machine learning: Progress in the machine intelligence for data science. *Sustainable Energy Technologies and Assessments*, *53*, 102557.
13. Biradar, S. H., Gorabal, J. V., & Gupta, G. (2022). Machine learning tool for exploring sentiment analysis on twitter data. Materials Today: Proceedings, 56, 1927-1934.
14. Villegas-Ch, W., Molina, S., Janón, V. D., Montalvo, E., & Mera-Navarrete, A. (2022, August). Proposal of a Method for the Analysis of Sentiments in Social Networks with the Use of R. In Informatics (Vol. 9, No. 3, p. 63). MDPI.
15. Jain, D. K., Boyapati, P., Venkatesh, J., & Prakash, M. (2022). An intelligent cognitive-inspired computing with big data analytics framework for sentiment analysis and classification. Information Processing & Management, 59(1), 102758.
16. Ali, F., Kwak, D., Khan, P., Islam, S. R., Kim, K. H., & Kwak, K. S. (2017).

Engineering Manufacture, 217(6), 779-790.
29. Nariman-Zadeh, N., DARVIZEH, A., & DARVIZEH, M. (2001). GMDH-Type Neural Network Modelling of Explosive Welding Process of Plates Using Singular Value Decomposition.
30. Nariman-Zadeh, N., Darvizeh, A., Darvizeh, M., & Gharababaei, H. (2002). Modelling of explosive cutting process of plates using GMDH-type neural network and singular value decomposition. Journal of Materials Processing Technology, 128(1-3), 80-87.
31. Mahendhiran, P. D., & Subramanian, K. (2022). CLSA-CapsNet: Dependency based concept level sentiment analysis for text. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-17.
32. Mandloi, L., & Patel, R. (2020, June). Twitter sentiments analysis using machine learninig methods. In 2020 International Conference for Emerging Technology (INCET) (pp. 1-5). IEEE.
33. Pathak, A. R., Pandey, M., & Rautaray, S. (2021). Topic-level sentiment analysis of social media data using deep learning. Applied Soft Computing, 108, 107440.
34. Revathy, G., Alghamdi, S. A., Alahmari, S. M., Yonbawi, S. R., Kumar, A., & Haq, M. A. (2022). Sentiment analysis using machine learning: Progress in the machine intelligence for data science. Sustainable Energy Technologies and Assessments, 53, 102557.
35. Singh, N. K., Tomar, D. S., & Sangaiah, A. K. (2020). Sentiment analysis: a review and comparative analysis over social media. Journal of Ambient Intelligence and Humanized Computing, 11(1), 97-117.
36. Villegas-Ch, W., Molina, S., Janón, V. D., Montalvo, E., & Mera-Navarrete, A. (2022, August). Proposal of a Method for the Analysis of Sentiments in Social Networks with the Use of R. In Informatics (Vol. 9, No. 3, p. 63). MDPI.
Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling. Transportation Research Part C: Emerging Technologies, 77, 33-48.
17. Alarifi, A., Tolba, A., Al-Makhadmeh, Z., & Said, W. (2020). A big data approach to sentiment analysis using greedy feature selection with cat swarm optimization-based long short-term memory neural networks. The Journal of Supercomputing, 76(6), 4414-4429.
18.https://www.kaggle.com/datasets/kazanova/sentiment140.
19.https://www.kaggle.com/datasets/techykajal/fakereal-news.
20.https://www.kaggle.com/datasets/elvinagammed/covid19-fake-news-dataset-nlp.
21.https://www.kaggle.com/datasets/marklvl/sentiment-labelled-sentences-data-set.
22.https://www.kaggle.com/datasets/mdepak/fakenewsnet.
23. Guo, H., Li, S., Qi, K., Guo, Y., & Xu, Z. (2018). Learning automata-based competition scheme to train deep neural networks. IEEE Transactions on Emerging Topics in Computational Intelligence, 4(2), 151-158.
24. Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. International journal of data mining & knowledge management process, 5(2), 1.
25. Gilpin, W. (2019). Cellular automata as convolutional neural networks. Physical Review E, 100(3), 032402.
26. Ivakhnenko, A. G. (1971). Polynomial theory of complex systems. IEEE transactions on Systems, Man, and Cybernetics, (4), 364-378.
27. Farlow, S. J. (1984). Self-Organizing Method in Modeling: GMDH. Type Algorithm.
28. Nariman-Zadeh, N., Darvizeh, A., & Ahmad-Zadeh, G. R. (2003). Hybrid genetic design of GMDH-type neural networks using singular value decomposition for modelling and prediction of the explosive cutting process. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of

37. Yoo, S., Song, J., & Jeong, O. (2018). Social media contents-based sentiment analysis and prediction system. Expert Systems with Applications, 105, 102-111.
38. Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. Knowledge and Information Systems, 60(2), 617-663.
39. Zong, C., Xia, R., & Zhang, J. (2021). Sentiment analysis and opinion mining. In Text Data Mining (pp. 163-199). Springer, Singapore.
40. Maity, D., Kanakaraddi, S., & Giraddi, S. (2023). Text Sentiment Analysis based on Multichannel Convolutional Neural Networks and Syntactic Structure. Procedia Computer Science, 218, 220-226.
41. Sodhar, I. N., Sulaiman, S., Buller, A. H., & Sodhar, A. N. (2023). Hybrid Approach Used to Analyze the Sentiments of Romanized Text (Sindhi). International Journal of Advanced Computer Science and Applications, 14(3).
42. Fazal, U., Khan, M., Maqbool, M. S., Bibi, H., & Nazeer, R. (2023). Sentiment Analysis of Omicron Tweets by using Machine Learning Models.