صفحات ۱۹ تا ۲۸

پردازش مفهوم زبانی پیام‌های کوتاه توییتر با استفاده از شبکه عصبی عمیق CNN

زهرا دره شوری [۱]، حسین مومن‌زاده [۲]، حسن ارفعی نیا [۳]

[۱] دانشگاه غیرانتفاعی لیان بوشهر ،
آدرس پست الکترونیکی
[۲] دانشگاه غیرانتفاعی لیان بوشهر
آدرس پست الکترونیکی
[۳] دانشگاه غیرانتفاعی لیان بوشهر
آدرس پست الکترونیکی

چکیده

تجزیه‌و تحلیل احساسات در رسانه های اجتماعی مانند توییتر (شبکه **X**) بسیار مهم و چالش برانگیز است. با توجه به ویژگی‌های چنین داده‌هایی نظیر طول توییت، اشتباهات املایی، اختصارات و کاراکترهای خاص، تجزیه‌و تحلیل احساسات نیازمند رویکردی غیرسنتی است. علاوه بر این، تجزیه و تحلیل احساسات رسانه های اجتماعی یک مسئله اساسی با کاربردهای جالب مانند هوش تجاری، نظارت پزشکی و امنیت ملی است. بیشتر روش‌های طبقه‌بندی احساسات رسانه‌های اجتماعی، قطبیت احساسات را بر اساس محتوای متنی قضاوت می‌کنند و از اطلاعات دیگر در این پلتفرم‌ها غفلت می‌کنند. در این تحقیق، ما چارچوب‌های مبتنی بر یادگیری عمیق را پیشنهاد می‌کنیم که اطلاعات رفتاری کاربر را نیز در درنظر می‌گیرد. در این مقاله، از این مدل‌های یادگیری چند وجهی **CNN** برای انجام تجزیه و تحلیل احساسات استفاده می‌کنیم. نتایج بدست‌آمده نشان می‌دهد که فراتر رفتن از محتوای یک سند در طبقه‌بندی احساسات سودمند است، زیرا درک عمیق‌تری از کار را در اختیار طبقه‌بندی‌کننده قرار می‌دهد و از لحاظ معیارهای دقت، صحت، حساسیت و **F1** بهبود قابل توجهی گزارش می‌شود.

کلمات کلیدی : پیام‌های کوتاه توییتر، شبکه های عصبی عمیق، مفهوم زبانی

۱- مقدمه

ظهور پلتفرم های رسانه های اجتماعی به کاربران وب فضایی برای بیان و به اشتراک گذاری افکار و نظرات خود در مورد انواع موضوعات و رویدادها داده است. یکی از محبوب‌ترین شبکه‌های اجتماعی X (Twitter سابق) است. این شبکه به افراد اجازه می دهد پیام هایی را برای بیان علایق، علاقه مندی ها، نظرات و احساسات خود نسبت به موضوعات و مسائل مختلفی که در زندگی روزمره با آن مواجه می شوند منتشر کنند. این پیام ها توییت نامیده می شوند که بلادرنگ و حداکثر ۱۴۰ کاراکتر هستند. توییتر حدود ۲۰۰ میلیارد توییت در سال، ۵۰۰ میلیون توییت در روز، ۳۵۰۰۰۰ توییت در دقیقه و ۶۰۰۰ توییت در ثانیه دارد [۱]. توییتر امکان دسترسی به نماهای ناخواسته مجموعه وسیعی از کاربران در مورد محصولات یا رویدادهای خاص را فراهم می کند. نظرات یا ابراز احساسات در مورد سازمان ها، محصولات، رویدادها و افراد برای بازاریابی [۲] و مطالعات اجتماعی [۳] بسیار مفید است. تجزیه و تحلیل احساسات در نظارت بر رسانه های اجتماعی ارزشمند است زیرا به فرد اجازه می دهد تا بینش هایی در زمینه ها یا موضوعات خاص به دست آورد و در مجموعه های متنوعی از زمینه ها مانند بازاریابی و تبلیغات، رسانه های اجتماعی، اقتصاد و علوم سیاسی استفاده می شود. پلتفرم‌های رسانه‌های اجتماعی مانند توییتر، فیس‌بوک و یوتیوب در حال حاضر از محبوب‌ترین مکان‌ها برای بحث و بررسی محصولات و خدمات جدید در بازارهای مختلف برای مشتریان هستند. با این حال، ماهیت آشفته ذاتی محتوای رسانه‌های اجتماعی، چالش‌های جدی را برای کاربردهای عملی تحلیل احساسات، مانند استخراج بازخورد معنی‌دار برای محصولات یا خدمات، درک کیفیت محصول، یا برآورده کردن انتظارات مشتری ایجاد می‌کند [۴].

برخلاف متون استاندارد با کلمات زیاد که به جمع آوری آمار کافی کمک می کند، متون در رسانه های اجتماعی، به ویژه توییتر، فقط از تعداد محدودی کاراکتر تشکیل شده است. علاوه بر این، هنگامی که یک کاربر پیامی (توئیت) ارسال می کند، ممکن است دارای اختصارات یا کلمات اختصاری جدیدی باشد که به ندرت در اسناد متنی معمولی ظاهر می شوند. بنابراین، اعمال روش های سنتی در چنین محیطی عملکرد قابل قبولی را به همراه نخواهد داشت. علاوه بر این، اکثر رویکردهای موجود برای تجزیه و تحلیل احساسات در میکروبلاگ ها، مانند توییتر، متون نوشته شده توسط کاربران مختلف را به طور یکسان در نظر می گیرند و ویژگی های کاربر را کنار می گذارند [۵].

چنین رویکردهایی تحلیل احساسات را به عنوان یک پدیده زبانی، بدون تأکید بر جنبه های نویسنده (کاربر) و جنبه های خاص میکروبلاگ، که از جنبه های غیر محتوایی بیرون می آیند، انجام می دهند. برای مثال، ویژگی‌های کاربر شامل حالات عاطفی (خلق) و سبک‌های نوشتاری است. جنبه های خاص میکروبلاگ شامل فعالیت اجتماعی کاربر، تراکم شبکه، کمی سازی احساسات موضوع و غیره است. در این مقاله ما به تجزیه و تحلیل احساسات در میکروبلاگ ها به عنوان وظیفه ای برای یکپارچه سازی اطلاعات از ابعاد مختلف، نه تنها از یک بعد (یعنی محتوا) نگاه می کنیم.

به طور کلی، سه روش اصلی برای تجزیه و تحلیل احساسات وجود دارد: (۱) روش مبتنی بر یادگیری ماشین (رویکرد نظارت شده)، (۲) روش مبتنی بر واژگان (رویکرد بدون نظارت)، و (۳) روش ترکیبی [۶]. اکثر مقالاتی که در مورد تحلیل احساسات محتوای رسانه‌های اجتماعی نوشته شده‌اند از رویکرد یادگیری ماشینی استفاده می‌کنند که در آن یک مجموعه داده برچسب‌گذاری شده برای آموزش یک الگوریتم استاندارد استفاده می‌شود [۴ و ۷-۱۰]. اشکال عمده این رویکرد توسعه مجموعه داده های مناسب است، زیرا طبقه بندی کننده های آموزش دیده عمدتاً وابسته به داده‌های برچسب گذاری شده ورودی هستند. الگوریتم‌های اصلی یادگیری ماشینی که در وظایف طبقه‌بندی احساسات Naive Bayes (NB) [۸] با دقت ۷۶ درصد و بردارهای ماشین پشتیبان(SVM) [۹] با دقت ۵۸ درصد هستند. J48 الگوریتمی برای تولید درخت تصمیم است که در مرجع [۱۰] پیشنهاد شده است. الگوریتم KNN، k نزدیک‌ترین همسایه‌ها را در بین نمونه‌های آموزشی برچسب‌گذاری‌شده بر اساس برخی معیارهای شباهت جستجو می‌کند، و آن k همسایه‌ها را بر اساس امتیاز شباهت‌ها فهرست می‌کند [۴]. در روش مبتنی بر واژگان، از یک یا چند فرهنگ لغت حاوی واژگان عاطفی برای تخمین محتوای عاطفی بخش‌های متن استفاده می‌شود. اینها فهرستهای کلماتی هستند که در آنها به هر کلمه یک ارزش عاطفی اختصاص داده شده است، به عنوان مثال سطح مثبت یا منفی که معمولاً منتقل می کند [۱۱]. ذکر این نکته ضروری است که این رویکرد قواعد مختلف نثر و نحوی را با هم ترکیب می کند تا دقت آن را افزایش دهد [۱۲]. تعداد قابل توجهی از فرهنگ لغت وجود دارد که به صورت خودکار یا نیمه خودکار توسعه یافته اند، به عنوان مثال، General Inquirer (GI)، SentiWordNet و WordNet-Affect. با این حال، اثربخشی رویکرد احساسات مبتنی بر واژگان با پوشش و دقت فرهنگ لغت مشخص می شود [۱۳]. همچنین در بسیاری از موارد تکیه بر کلمات عاطفی اغلب ناکافی است و به نتایج رضایت بخشی در تشخیص احساسات منجر نمی شود [۱۴]. روش ترکیبی، ترکیبی از یادگیری ماشین و رویکردهای مبتنی بر واژگان است. متن در ابتدا با استفاده از رویکرد مبتنی بر واژگان تحلیل می‌شود و خروجی تولید شده به عنوان داده‌های آموزشی به الگوریتم یادگیری ماشینی وارد می‌شود. خروجی فاز دوم متعاقباً برای گسترش فرهنگ لغت عاطفی استفاده می شود. کل فرآیند تا زمانی که برخی از معیارهای خاتمه برآورده شود تکرار می شود [۱۵]. یادگیری عمیق توسط بسیاری از محققین برای تحلیل احساسات توییتر بررسی و اجرا شده است. یکی از مهمترین روشها، مدل های LSTM است که در طبقه بندی احساسات به عملکرد چشمگیری دست یافته اند [۳]. آنها توانایی مدیریت وابستگی های دوربرد را دارند. در مرجع [۱۶]، رویکردی برای یادگیری کلمات خاص احساسات از توییت های جمع آوری شده با استفاده از نظارت پیشنهاد شده است. سه شبکه عصبی برای یادگیری توسعه یافته است. سپس از کلمات آموخته شده برای ایجاد ویژگی هایی برای طبقه بندی احساسات توییتر استفاده می شود. نویسندگان استدلال می کنند که عملکرد مناسب رویکرد آنها به این دلیل

حاصل می شود که شبکه اطلاعات احساسی متن را به جای بافت نحوی کلمات بدست می‌آورد. در مرجع [۱۷] از یک رویکرد یادگیری عمیق برای طبقه‌بندی توییت‌ها در هر دو سطح سند و عبارت، بر اساس شبکه‌های عصبی کانولوشنال استفاده کردند. در این مقاله از مدل شبکه عصبی برای مقداردهی اولیه کلمه embeddings استفاده می شود. این مدل بر روی مجموعه بزرگی از توییت ها آموزش داده شده است. مرحله دوم شامل اصلاح واژه embeddings است. سپس از یک شبکه عصبی کانولوشنال CNN و یک مجموعه بزرگ نظارت شده از راه دور استفاده می کنند. در نهایت از تعبیه‌های کلمه ایجاد شده و پارامترهای شبکه به‌دست‌آمده از مرحله دوم برای مقداردهی اولیه مدل استفاده می‌شود.

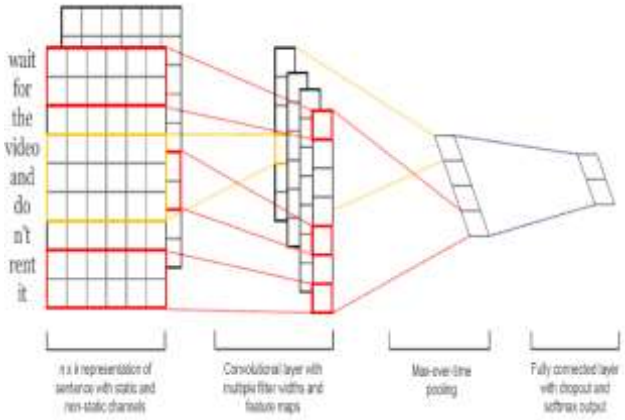
در این مقاله، می‌خواهیم، پیام های توییتر و هشتگ‌های مختلف را نیز در الگوریتم CNN وارد کنیم و از چندین مدل مختلف مبتنی بر CNN برای شناسایی یک مدل یادگیری عمیق دقیق برای تحلیل احساسات و رسیدن به جواب بهتر بهره گیریم. این نوع رویکرد عملکرد و دقت وظایف طبقه بندی احساسات را در رسانه های اجتماعی افزایش می دهد.

در بخش دوم مقاله، مدل یادگیری عمیق CNN را بررسی می‌کنیم. در بخش سوم الگوریتم پیشنهادی با استفاده تکنیک های یادگیری عمیق ارائه می‌کنیم. در بخش چهارم نیز به ارزیابی عملکرد روش پیشنهادی خود خواهیم پرداخت. نتیجه گیری در بخش پنجم ارائه می‌شوند.

۲- شبکه عصبی کانولوشنال (CNN)

CNN یک شبکه عصبی عمیق است که از لایه‌هایی با فیلترهای کانولوشن استفاده می‌کند که روی مجموعه‌ای از ویژگی‌ها اعمال می‌شوند [۱۸]. یک شبکه عصبی کانولوشن از یک لایه ورودی، لایه های پنهان و یک لایه خروجی تشکیل شده است. در یک شبکه عصبی کانولوشن، لایه‌های پنهان شامل لایه‌هایی هستند که کانولوشن را انجام می‌دهند. معمولاً این شبکه شامل یک لایه است که حاصل ضرب نقطه ای هسته کانولوشن را با ماتریس ورودی لایه انجام می دهد. همانطور که هسته کانولوشن در طول ماتریس ورودی لایه می لغزد، عملیات کانولوشن یک نقشه ویژگی ایجاد می کند که به نوبه خود به ورودی لایه بعدی کمک می کند. به دنبال آن لایه‌های دیگری مانند لایه‌های pooling، لایه‌های کاملاً متصل و لایه‌های نرمال‌سازی قرار می‌گیرند [۱۹].

از لایه جدول جستجو به منظور ایجاد ویژگی های محلی در اطراف هر کلمه جمله استفاده می شود. این ویژگی‌ها از لایه‌های کانولوشنال عبور می‌کنند. سپس ویژگی‌ها را در یک بردار ویژگی کلی ترکیب می‌کند که می‌تواند به لایه‌های کاملاً متصل تغذیه شود. این مدل در شکل ۱ نشان داده شده است. در این شکل مدل CNN دارای یک لایه کانوولوشن است که در آن ویژگی‌ها به جای جملات کامل از پنجره کوچکی از کلمات استخراج می‌شوند.

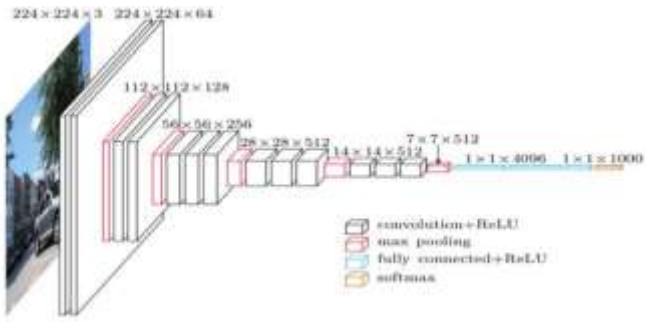


شکل ۱: مدل CNN برای تجزیه و تحلیل کلمات

۲-۱- مدل‌های مرجع CNN

۲-۱-۱- شبکه VGG 16

VGGNet یک معماری از شبکه CNN است که از بلوک‌های متشکل از تعداد افزایشی لایه‌های کانولوشن با فیلترهای اندازه ۳×۳ استفاده می‌کند (شکل ۲). بلوک‌های جمع‌آوری حداکثر بین بلوک‌های کانولوشن پراکنده شده‌اند و اندازه این نقشه‌های فعال‌سازی را به نصف کاهش می‌دهند. در نهایت از یک بلوک طبقه بندی استفاده شده است که شامل دو لایه متراکم ۴۰۹۶ نورون و آخرین لایه که لایه خروجی است از ۱۰۰۰ نورون تشکیل شده است [۲۰].



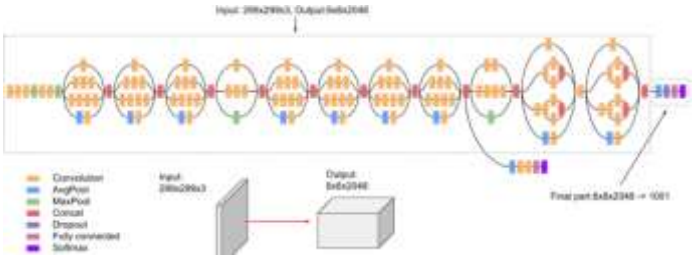
شکل (۲): ساختار معماری VGGNet

۱۶ و ۱۹ به تعداد لایه های وزنی که هر شبکه دارد، اشاره دارد. مشکل VGGNet این است که از ۱۳۸ میلیون پارامتر تشکیل شده است که ۳۴,۵ برابر بیشتر از GoogleNet است که اجرای آن را در کامپیوترهای معمولی چالش برانگیز می کند [۲۱].

۲-۱-۲- شبکه Inception v3

شبکه Inception v3 در مقایسه با VGGNet و شبکه‌های Inception (GoogLeNet/Inception v1) هم از نظر محاسباتی کارآمدتر است، هم از نظر تعداد پارامترهای تولید شده توسط شبکه و هم از لحاظ سخت‌افزار استفاده شده (حافظه و سایر منابع). اگر قرار است تغییراتی در شبکه آغازین ایجاد شود، باید مراقب باشید که مزایای محاسباتی از بین نرود [۲۳].

بنابراین، انطباق یک شبکه Inception برای موارد استفاده مختلف به دلیل عدم اطمینان از کارایی شبکه جدید، یک مشکل است. در یک مدل Inception v3، چندین تکنیک برای بهینه‌سازی شبکه پیشنهاد شده است تا محدودیت‌ها را برای تطبیق آسان‌تر مدل کاهش دهند. این تکنیک ها شامل پیچیدگی های فاکتوریزه، منظم سازی، کاهش ابعاد و محاسبات موازی شده است (شکل ۳). تعداد پارامترهای آن، ۵ میلیون برای شبکه (V1) و ۲۳ میلیون برای شبکه V3 است.



شکل (۳): ساختار شبکه Inception v3

۳- روش پیشنهادی مبتنی بر شبکه CNN

برخلاف متون استاندارد با کلمات زیاد که به جمع آوری آماره کافی کمک می‌کند، متون در شبکه توییتر، فقط از تعداد محدودی کاراکتر تشکیل شده است. علاوه بر این، هنگامی که یک کاربر توئیتی را ارسال می کند، ممکن است دارای کلمات اختصاری جدیدی باشد که به ندرت در اسناد متنی معمولی ظاهر می شوند. بنابراین، اعمال روش های سنتی در چنین محیطی عملکرد قابل قبولی را به همراه نخواهد داشت. برای پرداختن به این موضوع، ما یک مدل مبتنی بر شبکه عصبی CNN را پیشنهاد می‌کنیم که ویژگی های متفاوت توییت و کاربر را درنظر می‌گیرد.

۳-۱- جاسازی کلمات

در سیستم های پردازش زبان طبیعی، کلمات به صورت نمادهای گسسته بیان می شوند. از آنجایی که این قالب رمزگذاری هر کلمه را به طور مستقل مورد بررسی قرار می دهد، هنگام تلاش برای شناسایی کلمات مشابه معنایی، اطلاعات اضافی را در اختیار سیستم قرار نمی دهد [۲۵]. جاسازی کلمه ها شامل فهرست‌هایی از کلمات w در یک فضای nبعدی است که در آن کلمات از نظر معنایی مشابه همسایه‌هایی هستند که می‌توانند با استفاده از فرهنگ لغت تولید شوند. اگر $T \in \{w_1, w_2, ..., w_n\}$ یک متن با n کلمه باشد که بر اساس متن متفاوت خواهد بود و $f(w_i) = v_i$ یک تابع نگاشت باشد که کلمه w_i را در بردار v_i ترسیم می کند. روش های جاسازی کلمه به طور کلی به دو دسته، شامل تعبیه مبتنی بر فرکانس و مبتنی بر پیش بینی طبقه بندی می شوند [۲۶]. GloVe یک ابزار موثر مبتنی بر فرکانس است که از یک الگوریتم یادگیری بدون نظارت برای به دست آوردن نمایش های برداری برای

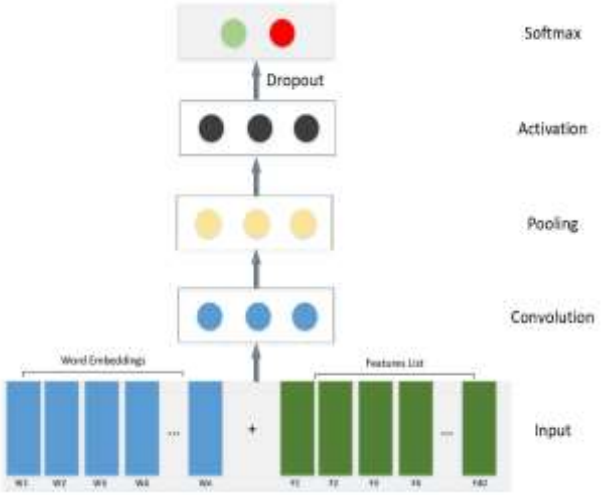
کلمات اســتفاده می کند. word2vec یکی از مدل های پیش بینی کارآمد محاســباتی برای یادگیری جاســازی کلمات از NLP اســت. word2vec از مجموعه کلمات پیوســته (CBOW) یا مدل Skip-gram [۲۶] اســتفاده می کند. در این مقاله از ما از روش word2vec اســتفاده کرده‌ایم. به منظور ساخت بردارهای تعبیه word2vec از طریق توییت های داده شده، از بسته Genism از پایتون استفاده کردیم. جدول ۱ فهرست پارامترهای مورد استفاده در مدل word2vec را شرح می دهد:

جدول ۱: پارامترهای استفاده شده در مدل Word2vec

Parameter name	Value(s)
Dimensionality of the word vectors (size)	100-300
Maximum distance between the current and predicted word within the distance (window)	5-15
Total frequency threshold (min_count)	5-15
No of worker threads to train the model (workers)	8-10
Training iterations (epochs)	16-50
Maximum padding length	150-300

۳-۲- مدل پیشنهادی CNN

مدل پیشــنهادی ما برای تحلیل احســاســات از یک شــبکه عصبی کانولوشنال تشکیل شده است. معماری سیستم در شکل ۴ ارائه شده است. این مدل با استفاده از کتابخانه Weka پیاده سازی شده است.



شکل۴: ساختار مدل پیشنهادی CNN

اجزای اصــلی شــبکه عبارتند از: لایه های ورودی، کانولوشــن، ادغام، فعال سازی و softmax. لایه ورودی شامل جاسازی های کلمه و لیستی از ویژگی ها اســت. Word Embedding به طور تصــادفی مقداردهی اولیه شــده است. علاوه بر این، ویژگی‌هایی که نویسنده توییت را توصیف می‌کنند به بردار تولید شــده اضــافه می‌شــوند و سپس به لایه بعدی وارد می‌شوند.در لایه کانولوشن، هدف اصلی استخراج الگوها است. به منظور امکان یادگیری مرزهای تصمیم گیری غیر خطی، یک تابع فعال ســازی غیر خطی در لایه فعال ســازی قرار دارد. برای تابع فعال سازی از ReLU استفاده می کنیم زیرا سرعت آموزش را افزایش می دهد و نتایج دقیق تری نسبت به سایر توابع فعال سازی ایجاد می کند. مقادیر پارامترهای شبکه در جدول ۲ آورده شده است.

جدول ۲: مقادیر پارامترهای شبکه CNN

Parameter name	Value(s)
Maximum number of words (restricted for the same length)	30-42
Number of classes for output layer	2
Size of the vocabulary	25,000-30,000
Output dimensionality	140-200
Size of the kernel	3-8
Activation functions	relu, tanh
Batch size	16, 32,64
Pooling types	2, maximum
Number of epochs (number of cycles)	20-50
Dropout rate probability	0.2-0.5
Optimizer	adagrad, adam
Loss function	Categorical crossentropy
Number hidden nodes	16-64

جدول ۳: لیست ویژگی های مورد استفاده

F1 تعداد توییت های یک کاربر
{F2، F3، F4} تعداد توییت‌های {مثبت، منفی، خنثی} ارسال شده توسط یک کاربر.
{F5، F6، F7} احتمال نگرش {مثبت، منفی، خنثی} یک کاربر.
{F8,F9,F10}میانگین تعداد توییت‌های {مثبت، منفی، خنثی} ارسال شده توسط یک کاربر.
F11 تعداد فالوورها.
F12 تعداد دوستان.
F13 تایید حساب.
F14، F15 تعداد صفت ها و میانگین آنها.
F16، F17 تعداد اسمها و میانگین آنها.
F18، F19 تعداد قیدها و میانگین آنها.
F20، F21 تعداد افعال و میانگین آنها.
F22، F23 تعداد هشتگ ها و میانگین آنها.
F24، F25 تعداد اشاره‌ها (mention) و میانگین آنها.
F26، F27 تعداد URL ها و میانگین آنها.
F28، F29 تعداد شکلک ها و میانگین آنها.
F30، F31 تعداد علامت سوال و میانگین آنها.
F32، F33 تعداد علامت تعجب و میانگین آنها.
F34، F35 تعداد کلمات در هر توییت و میانگین آنها.
F36، F73 تعداد کلمات مثبت در واژگان بینگ لیو و میانگین آنها.
F38، F39 تعداد کلمات منفی در واژگان بینگ لیو و میانگین آنها.

جزئیات تمام ۴۰ ویژگی مورد استفاده برای آموزش مدل پیشنهادی در جدول ۳ آورده شده است. این ویژگی‌ها از مجموعه SemEval-2016 استخراج شده است.

۴- پیاده‌سازی و بررسی نتایج بدست‌آمده

ما مدل خود را بر روی دو مجموعه داده آموزش می دهیم که مجموعه داده های توییتر هستند که توسط SemEval-2016 منتشر شده اند. مجموعه داده اول (SemEval-2016_1) از ۷۷۴۲۴۴ توییت و دومین مجموعه (SemEval-2016_2) از ۴۹۱۹۰۲ توییت تشکیل شده است. این دو مجموعه داده دارای شناسه توییت همراه با حاشیه‌نویسی با برچسب‌های مثبت و منفی هستند.

۴-۱- پیش پردازش داده ها

توییت‌های خام دارای داده‌های بی‌معنی و بدون ساختار و کلمات تکراری هستند. از این رو، هر آزمایش چند مرحله پیش پردازش را دنبال کرد تا تغییراتی در متن توییت‌ها ایجاد شود تا تجزیه و تحلیل متن بیشتر انجام شود. مجموعه داده‌ها مراحل پیش‌پردازش زیر را طی کردند:

۱) تبدیل متن به حروف کوچک

۲) حذف تمام ردیف های پوچ

۳) حذف تمام فضاهای خالی اضافی

۴) حذف تمام اعداد، حروف عددی و نقطه گذاری

۴-۲- گروه‌بندی مجموعه‌های ویژگی‌ها

از آنجایی که بسیاری از ویژگی‌ها در آزمایش‌های ما گنجانده شده‌اند، آن‌ها را در مجموعه‌ها گروه‌بندی کردیم. علاوه بر این، استفاده از مجموعه ها وضوح بیشتری را برای ما فراهم می کند که کدام مجموعه تأثیر بیشتری بر عملکرد مدل دارد. این مجموعه ها در جدول ۴ نشان داده شده است:

جدول ۴: استفاده از ویژگی‌ها در ترکیب های مختلف با دیتاست

Feature Set ID	Features Used
Set No. 1	F5, F6, and F7.
Set No. 2	F1 to F13.
Set No. 3	F2, F3, and F4.
Set No. 4	Word Embedding, F5, F6, and F7.
Set No. 5	Word Embedding.
Set No. 6	All features (F1 to F40).
Set No. 7	All features (F1 to F40) + Word Embedding.

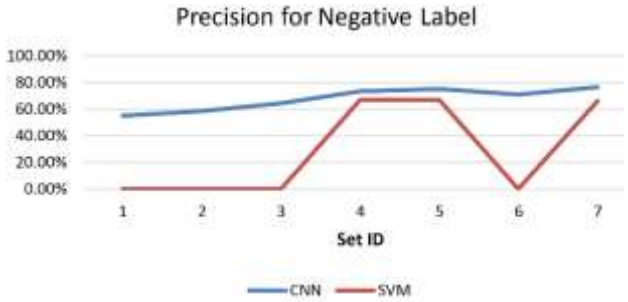
۴-۳- نتایج بدست‌آمده

در اینجا ما یک ارزیابی عملکرد مقایسه ای از هر مدل از نظر پیش بینی صحیح ارائه می کنیم. نتایج جدول ۵ برای صحت، یادآوری، دقت، و F1 برای شش الگوریتم به دست آمده است.

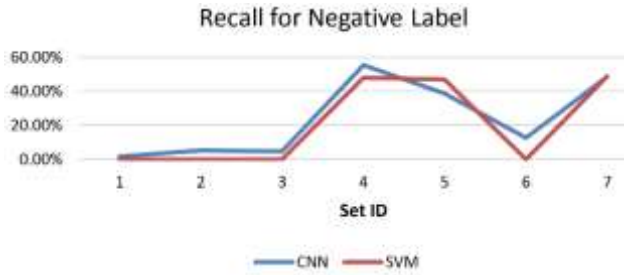
به طور کلی، طبقه‌بندی‌کننده‌های یادگیری عمیق (CNN و LSTM) عملکرد خود را برخلاف سایر طبقه‌بندی‌کننده‌ها در یک روند ثابت نگه می‌دارند. بهترین دقت ۸۸,۷۱ درصد برای روش پیشنهادی CNN در مجموعه شماره ۴ بود، پس از آن LSTM با ۸۸,۱۳ درصد در همان مجموعه، و کمترین مقدار برای NB در مجموعه شماره ۲ ۴۸,۳۱ درصد بود. تأثیر مجموعه داده نامتعادل ۷ به وضوح بر روی NB و SVN مشاهده می‌شود. یکی از دلایل احتمالی که مجموعه داده نامتعادل تأثیر قابل توجهی بر عملکرد طبقه‌بندی‌کننده CNN پیشنهادی ندارد، نحوه محاسبه وزن‌ها در این نوع طبقه‌بندی‌کننده است. می توان مشاهده کرد که بهترین دقت عملکرد طبقه بندی توسط CNN (پیشنهادی)، LSTM و SVM به دست می‌آید. با این حال، همانطور که در شکل ۵ و ۶ نشان داده شده است، CNN دقت و یادآوری بالاتری نسبت به SVM در تمام مجموعه ها دارد.

جدول ۵: نتایج بدست آمده میزان دقت روش CNN (پیشنهادی)، SVM، NB، J48، KNN و LSTM

Set No.	1	2	3	4	5	6	7
CNN	82.63%	82.85%	82.93%	88.71%	87.08%	83.90%	88.46%
SVM	82.58%	82.58%	82.58%	86.84%	86.67%	82.58%	86.75%
NB	82.63%	48.31%	83.62%	85.16%	85.16%	59.38%	70.40%
J48	82.58%	82.25%	82.22%	87.25%	87.61%	82.33%	85.44%
KNN	81.78%	81.56%	80.96%	82.72%	82.66%	82.14%	82.83%
LSTM	83.63%	77.14%	84.29%	88.13%	84.83%	83.74%	86.48%



شکل ۵: معیاردقت برای برچسب‌های منفی



شکل ۶: معیاریادآوری (حساسیت) برای برچسب‌های منفی

این بدان معنی است که استفاده از دقت به عنوان تنها معیار اندازه گیری برای عملکرد طبقه بندی کننده کافی نیست. اگرچه تعداد نمونه های منفی در هر دو مجموعه داده کمتر از نمونه های مثبت است، مدل CNN پیشنهادی می‌تواند عملکرد خوب و پایداری را به ما ارائه دهد.

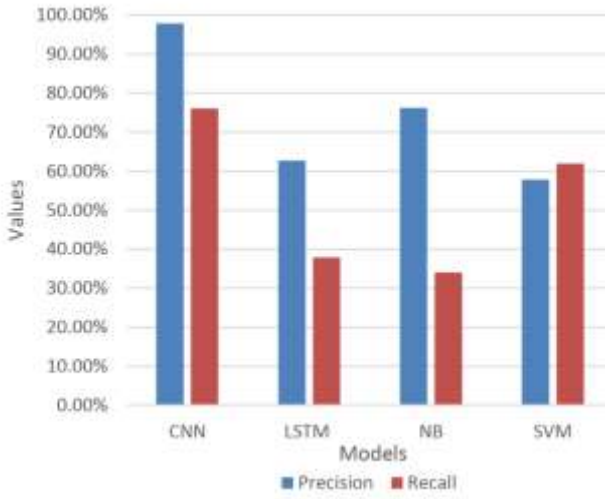
عملکرد ضعیف NB تعجب آور نیست زیرا بر فرض استقلال مشروط در میان ویژگی ها متکی است، که به وضوح در اینجا درست نیست. همه ویژگی‌های مورد استفاده در آزمایش‌ها درجاتی از وابستگی را در بین آنها دارند.

بر اساس مجموعه ها، بهترین دقت توسط مجموعه های شماره ۴ و ۷ به دست آمد. این انگیزه ما را برای استفاده از اطلاعات فراتر از محتوای یک توییت مشخص پشتیبانی می کند. برای بررسی نتایج این دو مجموعه برای تحلیل احساسات، F1 به همراه نمرات دقت و یادآوری شش مدل در هر دو مجموعه ۴ و ۷ در جدول ۶ آورده شده است. این معیارها بینش بیشتری در مورد عملکرد طبقه‌بندی‌کننده‌ها نسبت به متریک دقت به ما می‌دهند.

جدول ۶: امتیاز معیارهای F1، دقت و یادآوری شش مدل در مجموعه های شماره ۴ و ۷

	Precision		Recall		F1	
	Set No.		Set No.		Set No.	
	4	7	4	7	4	7
CNN	0.88	0.88	0.89	0.89	0.88	0.87
SVM	0.86	0.86	0.87	0.87	0.86	0.86
NB	0.85	0.83	0.85	0.70	0.85	0.74
J48	0.86	0.84	0.87	0.85	0.86	0.85
KNN	0.81	0.84	0.83	0.83	0.75	0.75
LSTM	0.87	0.86	0.88	0.86	0.86	0.83

در جدول ۶، بالاترین امتیازات F1 توسط مدل CNN پیشنهادی به دست آمده است (۰,۸۸ در مجموعه شماره ۴ و ۰,۸۷ در مجموعه شماره ۷). همچنین شایان ذکر است که مدل CNN عملکرد خوب ثابتی را در هر دو مجموعه نشان می‌دهد و به دنبال آن مدل LSTM قرار دارد. این نتایج تأیید می کند که استفاده از اطلاعات رفتاری کاربر علاوه بر محتوای متنی در تحلیل احساسات با بهبود دقت طبقه بندی می باشد.



شکل ۷: معیارهای یادآوری و دقت برای داده‌های برچسب مثبت

شـکل ۷ نتایج به‌دسـت‌آمده از روش‌های پایه و مدل پیشنهادی را در دقت و یادآوری نشان می‌دهد. در بین مدل های پایه، بالاترین دقت، ۸۸,۶۹ درصد، در مجموعه داده ۴ با استفاده از LSTM، و ۷۷,۷۸ درصد در مجموعه داده ۷ با استفاده از SVM به دست آمده است. از سوی دیگر، NB کمترین عملکرد را در دقت در هر دو مجموعه داده ارائه می دهد. این شکل نشان می‌دهد که مدل پیشنهادی ما با یک حاشیه قابل‌توجه از همه مدل‌ها بهتر عمل می‌کنند.

۵- نتیجه گیری

در این مقاله ما یک مدل تحلیل احسـاسـات را ارائه کردیم که با ترکیب لیسـتی از ویژگی‌ها توسـعه یافته است. ما معماری یک شبکه عصبی کانولوشنال (CNN) را پیشـنهاد کردیم که نه تنها توییت‌های کاربر، بلکه رفتار کاربر را نیز در نظر می‌گیرد. ارزیابی ما کارایی مدل را در یک محیط رسـانه اجتماعی نشـان داد. مدل ما در دقت، یادآوری، دقت و F1 از روش‌های پایه بهتر عمل کرد. علاوه بر این، مدل پیشنهادی کمتر تحت تأثیر مسائل داده ناهمگن قرار گرفـت. علاوه بر این، این رویکرد بر مسـئلـه نیـاز بـه مجموعه داده بزرگ برای آموزش مدل‌های یادگیری عمیق مانند LSTM غلبه کرد. نتایج تجربی اثربخشـی رویکرد پیشـنهادی را در مقایسـه با سـایر مدل‌های پایه که فقط از محتوای متنی اسـتفاده می‌کنند، نشان داد. به طور خاص، نتایج بهبود قابل توجهی را در عملکرد و دقت وظایف طبقه‌بندی احساسات با رویکرد جدید نشان داد.

مراجع

[۱] A. S. M. Alharbi and E. de Doncker, “Enhance a Deep Neural Network Model for Twitter Sentiment Analysis by Incorporating User Behavioral Information”, in, ser. Lecture Notes in Computer Science, D.-S. Huang, V. Bevilacqua, P. Premaratne, and P. Gupta, Eds., vol. 10361, Cham: Springer International Publishing, 2017, pp. 81–88, isbn: 978-3-319-63308-4. doi: 10.1007/978-3-319-63309-1_8. [Online].

[۲] A. S. M. Alharbi and E. de Doncker, “Emotional Awareness based Classification Model for Twitter Sentiment Analysis using a Deep Neural Network”, in *Proceedings of the 21st International Conference on Artificial Intelligence ICAI’19*, CSREA Press, 2019, pp. 142–145, isbn: 1-60132-501-0.

[۳] A. S. M. Alharbi and E. de Doncker, “Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information”, *Cognitive Systems Research*, vol. 54, pp. 50–61, 2019, issn: 13890417. doi: 10.1016/j.cogsys. 2018.10.001. [Online].

[۴] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, “Target-dependent Twitter Sentiment Classification”, *Computational Linguistics*, pp. 151–160, 2021. [Online]. Available: http://www.aclweb.org/anthology/P11-1016.

[۵] M. Thelwall, K. Buckley, and G. Paltoglou, “Sentiment in Twitter events”, *Journal of the American Society for Information Science and Technology*, vol. 62, no. 2, pp. 406– 418, 2021, issn: 15322882. doi: 10.1002/asi.21462. arXiv: 0912.3882.

[۶] J. Lin, W. Mao, and D. D. Zeng, “Personality-based refinement for sentiment classification in microblog”, *Knowledge-Based Systems*, vol. 132, pp. 204–214, 2019, issn: 09507051. doi: 10.1016/j.knosys.2017.06.031. [Online].

[۷] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, “Psychological Aspects of Natural Language Use: Our Words, Our Selves”, *Annual Review of Psychology*, vol. 54, no. 1, pp. 547–577, 2022, issn: 0066-4308. doi: 10.1146/annurev.psych.54. 101601.145041. [Online].

[۸] Social Networks”, *Annual Review of Sociology*, M. McPherson, L. Smith-Lovin, and J. M. Cook, “Birds of a Feather: Homophily in vol. 27, no. 1, pp. 415–444, 2001, issn: 0360-0572. doi: 10.1146/annurev.soc.27.1.415. arXiv: 9702163 [cond-mat].

[۹] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks", *Proceedings of the National Academy of Sciences*, vol. 111, no. 24, pp. 8788–8790, 2019, issn: 0027-8424. doi: 10.1073/pnas.1320040111.

[۱۰] X. Hu, L. Tang, J. Tang, and H. Liu, "Exploiting social relations for sentiment analysis in microblogging", in *Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13*, vol. 1, New York, New York, USA: ACM Press, 2018, p. 537, isbn: 9781450318693. doi: 10.1145/2433396.2433465.

[۱۱] [11] and Explicit Similarity Network for User-Level Sentiment Classification of Microblogging", in *PRICAI 2016: Trends in Artificial Intelligence: 14th Pacific Rim International Conference on Artificial Intelligence*, 2016, pp. 180–192, isbn: 978-3-319-42910-6. doi: 10.1007/978-3-319- 42911- 3. [Online]. Available: http://link.springer.com/10.1007/978-3-319-42911-3.

[۱۲] J. Read, "Using emoticons to reduce dependency in machine learning techniques for sentiment classification", in *Proceedings of the ACL Student Research Workshop on - ACL '05*, Morristown, NJ, USA: Association for Computational Linguistics, 2015, p. 43, isbn: 1932432515. doi: 10.3115/1628960.1628969. [Online].

[۱۳] L. Barbosa and J. Feng, "Robust Sentiment Detection on Twitter from Biased and Noisy Data", in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, Beijing, China, 2010, pp. 36–44, isbn: 1012341234. doi: 10. 1016 / j . sedgeo . 2006 . 07 . 004. arXiv: 1702 . 08388. [Online]. Available: http : //dl.acm.org/citation.cfm?id=1944566.1944571.

[۱۴] A. Bifet and E. Frank, "Sentiment Knowledge Discovery in Twitter Streaming Data", in *Proceedings of the 13th international conference on Discovery science*, Canberra, Australia, 2010, pp. 1–15, isbn: 978-3-642-16183-4. doi: 10.1007/978-3-642-16183- 4.

[۱۵] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", in *Seventh Conference on International Language Resources and Evaluation*, 2010, pp. 1320–1326.

[۱۶] J. Cigarrán, Á. Castellanos, and A. García-Serrano, "A step forward for Topic Detection in Twitter: An FCA-based Approach", *Expert Systems with Applications*, vol. 57, pp. 21–36, 2016, issn: 09574174. doi: 10.1016/j.eswa.2016.03.011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0957417416301038.

[۱۷] A. Giachanou and F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods", *ACM Computing Surveys*, vol. 49, no. 2, pp. 1–41, 2016, issn: 0360-0300. doi: 10.1145/2938640. [Online]. Available: http://dx.doi.org/10.1145/2938640.

[۱۸] C. Baziotis, N. Pelekis, and C. Doulkeridis, "DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis", in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval- 2017)*, Vancouver, Canada: Association for Computational Linguistics, 2017, pp. 747– 754. [Online]. Available: http://www.aclweb.org/anthology/S17-2126.

[۱۹] A. Bella, C. Ferri, J. Hernandez-Orallo, and M. J. Ramirez-Quintana, "Quantification via Probability Estimators", in *2010 IEEE International Conference on Data Mining*, IEEE, 2010, pp. 737–742, isbn: 978-1-4244-9131-5. doi: 10.1109/ICDM.2010.75. [Online]. Available: http://ieeexplore.ieee.org/document/5694031/.

[۲۰] G. Paltoglou and A. Giachanou, "Opinion Retrieval: Searching for Opinions in Social Media", in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8830, 2014, pp. 193– 214, isbn: 10.1007/978-3-319-12511-4_10. doi: 10.1007/978-3-319-12511-4_10. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84921407090{\&}partnerID=40{\&}md5=ab1923e2f9e3d1e308d0f2cdd664a480http: //link.springer.com/10.1007/978-3-319-12511-4{_}10.

[۲۱] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis", *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008, issn: 1554-0669. doi: 10.1561/1500000001. [Online]. Available: http://www.nowpublishers.com/ article/Details/INR-001.

[۲۲] B. Liu, Sentiment Analysis Mining Opinions, Sentiment, and Emotions. Cambridge University Press, 2015, p. 367, isbn: 978-1-107- 01789-4.

[۲۳] D. Das and S. Bandyopadhyay, "Emotion Analysis on Social Media: Natural Language Processing Approaches and Applications", *Spring-Verlag Wien*, pp. 19–37, 2014. doi: 10.1007/978-3-7091-1340-0. [Online]. Available: http://link.springer.com/ 10.1007/978-3-7091-1340-0.

[۲۴] D. Maynard and A. Funk, "Automatic detection of political opinions in tweets", *CEUR Workshop Proceedings*, vol. 718, pp. 81–92, 2011, issn: 16130073. doi: 10. 1007/978-3-642-25953-1_8.

[۲۵] S.-M. Kim, S.-M. Kim, E. Hovy, and E. Hovy, "Determining the sentiment of opinions", *Proceedings of the 20th international conference on Computational Linguistics - COLING '04*, 1367–es, 2014. doi: 10.3115/1220355.1220555. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1220355.1220555.

A. Muhammad, N. Wiratunga, and R. Lothian, "A Hybrid Sentiment Lexicon for Social Media Mining", in *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, IEEE, 2014, pp. 461–468, isbn: 978-1-4799-6572-4. doi: 10. 1109/ICTAI.2014.76. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/