مجله محاسبات نرم، جلد ۱۳، شماره ۱ (بهار و تابستان ۱۴۰۳)، صفحه ۹۶ـ۱۰۷
https://doi.org/10.22052/scj.2025.248730.1122

مجله محاسبات نرم
SOFT COMPUTING JOURNAL
تارنمای مجله: scj.kashanu.ac.ir

University of Kashan

بهبود فرآیند عقیده کاوی به کمک تلفیق روش‌های گرگ خاکستری و ماشین بردار پشتیبان✧

فریبا صلاحی[1]*، استادیار

[1] گروه مدیریت صنعتی، دانشکده مدیریت، دانشگاه آزاد اسلامی واحد الکترونیکی، تهران، ایران.

اطلاعات مقاله

تاریخچه مقاله:
دریافت ۴ دی ماه ۱۴۰۱
پذیرش ۳۱ شهریور ماه ۱۴۰۲

کلمات کلیدی:
عقیده کاوی
ماشین بردار پشتیبان
الگوریتم بهینه‌سازی گرگ خاکستری
تعیین ویژگی

چکیده

ظهور وب و رشد مستمر آن، حجم عظیمی از اطلاعات تولید شده توسط کاربران را ایجاد کرده است. در این اطلاعات، به راحتی می‌توان اطلاعات ذهنی ارزشمندی را پیدا کرد، به ویژه در شبکه‌های اجتماعی و بسترهای تجارت االکترونیک که حاوی اطلاعات مهمی در مورد کاربران هستند. حوزه عقیده کاوی در سال‌های گذشته به شدت مورد توجه قرار گرفته است. هر روز مقالات تحقیقاتی جدیدی منتشر می‌شوند که در آنها رویکردهای مختلف هوش مصنوعی برای وظایف و برنامه‌های مختلف مرتبط با عقیده کاوی اعمال می‌شوند. در این مقاله رویکرد جدیدی مبتنی بر ماشین بردار پشتیبان و الگوریتم بهینه‌سازی گرگ‌های خاکستری برای بهبود فرآیند عقیده کاوی پیشنهاد شده است، به گونه‌ای که الگوریتم گرگ‌های خاکستری برای تعیین ویژگی‌های موثر در فرآیند عقیده کاوی به کار رفته است و عملکرد ماشین بردار پشتیبان را بهبود می‌دهد. نتایج پیاده‌سازی این تحقیق نشان داد که سیستم پیشنهادی توانسته با انتخاب ویژگی‌های موثر به افزایش دقت و پوشاندن خطای رویکرد ماشین بردار پشتیبان کمک کند. سیستم پیشنهادی با استفاده از سه معیار صحت، فراخوانی و دقت مورد ارزیابی قرار گرفت، که صحت برای کلاس اول و دوم به ترتیب برابر با ۰/۶۸ و ۰/۹۲ درصد، فراخوانی برای کلاس اول و دوم به ترتیب برابر با ۰/۹۴ و ۰/۶۳ درصد و دقت ۰/۷۷ درصد بوده است. نتایج حاکی از آن است سیستم پیشنهادی این تحقیق توانسته در هر دو کلاس به نتایج مطلوبی دست پیدا کند.

۱. مقدمه

بخش مهمی از رفتار جمع‌آوری اطلاعات ما همیشه این بوده است که بفهمیم دیگران چه فکر می‌کنند. با در دسترس بودن و محبوبیت فزآینده منابع سرشار از عقاید فرصت‌ها و چالش‌های جدیدی به وجود می‌آیند که با استفاده از آنها می‌توان از فناوری‌های اطلاعات برای جستجو و درک نظرات دیگران استفاده کرد. فعالیت در حوزه نظرکاوی و تحلیل احساسات که با پردازش محاسباتی عقیده، احساسات و ذهنیت در متن سر و کار دارد، حداقل تا حدی به عنوان پاسخی مستقیم به افزایش علاقه به سیستم‌های جدید رخ داده است که به طور مستقیم با نظرات به عنوان یک شیء درجه یک سر و کار دارند [۱].

نظرکاوی، زیرشاخه‌ای در داده‌کاوی و زبان شناسی محاسباتی است که به رویکردهای محاسباتی برای استخراج، طبقه‌بندی، درک و ارزیابی نظرات بیان شده در منابع مختلف خبری برخط،

✧ نوع مقاله: پژوهشی
* نویسنده مسئول
پست(های) الکترونیک: salahi_en@yahoo.com (صالحی)

نحوه ارجاع به مقاله: فریبا صلاحی، «بهبود فرآیند عقیده کاوی به کمک تلفیق روش‌های گرگ خاکستری و ماشین بردار پشتیبان»، مجله محاسبات نرم، جلد ۱۳، شماره ۱، صفحات ۹۶ـ۱۰۷، بهار و تابستان ۱۴۰۳.

نظرات رسانه‌های اجتماعی و سایر محتوای تولید شده توسط کاربر اشاره دارد. تحلیل احساسات اغلب در نظرکاوی برای شناسایی احساسات، عاطفه، ذهنیت و سایر حالات احساسی در متن برخط استفاده می‌شود [۲]. عقیده‌ها که به صورت داده‌های متنی جمع‌آوری می‌شوند، دارای اطلاعات ارزشمندی هستند که با توجه زمینه کاربرد، می‌توانند اطلاعات مفیدی را برای افراد و سازمان‌ها تهیه کنند. عقیده‌کاوی به دنبال مشخص نمودن جهت عقیده‌ها یا به عبارتی دیگر، مثبت و منفی بودن هر عقیده است [۳]. طبقه‌بندی نظرات یا احساسات، مهمترین وظیفه در حوزه عقیده‌کاوی است. با توجه به اینکه در سال‌های اخیر، ثبت نظرات مردمی به صورت برخط به طور چشمگیری افزایش یافته است، به دست آوردن نتایج عقیده‌کاوی یکی از زمینه‌های اصلی محاسبات احساسات در نظر گرفته می‌شود [۴].

الگوریتم‌های متعددی در زمینه عقیده‌کاوی توسعه یافته است که هر یک در بهبود این مساله تلاش کرده‌اند. در این مقاله نیز هدف آن است که یکی از جدیدترین روش‌های فراابتکاری ارائه شده را به منظور عقیده‌کاوی بکار بگیریم. علت انتخاب این روش طبقه‌بندی، سادگی بکارگیری آن در حل مسائل درک نظرات می‌باشد. رویکرد جدید با ترکیب ماشین بردار پشتیبان و الگوریتم گرگ خاکستری ارائه گردیده که به دنبال عقیده‌کاوی و تجزیه و تحلیل احساسات موجود در متن‌ها می‌باشد و عقیده مثبت و منفی موجود در متن‌ها را مشخص می‌کند. انتخاب ویژگی با الگوریتم گرگ خاکستری سبب کاهش ابعاد و همچنین افزایش دقت می‌شود. این الگوریتم با هدف بهبود عملکرد جستجو، افزایش سرعت همگرایی و جلوگیری از گیر افتادن در بهینه محلی ارائه شده است.

۲. پیشینه تحقیق

در سال‌های اخیر، مشاهده شده است که عقیده‌کاوی در سطح وب و رسانه‌های اجتماعی به شکل‌دهی مجدد کسب و کارها و تحت تاثیر قرار دادن احساسات و عواطف عمومی کمک کرده است و عمیقا بر سیستم‌های اجتماعی و سیاسی ما تاثیر گذاشته است. لذا جمع‌آوری و مطالعه نظرات در وب به یک ضرورت تبدیل شده است.

کومار و همکارانش یک استراتژی انتخاب ویژگی مبتنی بر معنا برای انجام وظایف نظرکاوی معرفی کردند [۵]. مجموعه ویژگی‌ها، با کمک رویکرد مبتنی بر معنای معرفی شده و به منظور در نظر گرفتن توانایی پیش‌بینی فردی کلمات و به حداقل رساندن ویژگی‌های انتخاب شده، مورد بررسی قرار گرفتند.

گانش و همکاران از انواع طبقه‌بندی‌کننده‌ها مانند نایوبیز، جنگل تصادفی، درخت تصمیم، ماشین بردار پشتیبان، K-نزدیکترین همسایه و رگرسیون لجستیک برای طبقه‌بندی متن استفاده کردند [۶]. در این مقاله طبقه‌بندی‌کننده رگرسیون لجستیک برای طبقه‌بندی متن و ایجاد حس کلمه بالاترین صحت[1]، یادآوری و امتیاز F1 را در میان طبقه‌بندی‌کننده‌های دیگر به دست آورد. امتیاز F1 میانگین وزنی دقت و یادآوری است. این امتیاز به طور معمول مفیدتر از دقت است، به خصوص اگر توزیع کلاس ناهموار باشد. علاوه بر این، طبقه‌بندی‌کننده نایوبیز بهترین دقت[2] را در بین طبقه‌بندی‌کننده‌های مورد استفاده به دست آورد.

در مقاله آبدلمینام و همکاران، در بخش اول، از شش روش یادگیری ماشین از جمله درخت‌های تصمیم، رگرسیون لجستیک، K-نزدیک‌ترین همسایه، جنگل‌های تصادفی، ماشین‌های بردار پشتیبان و نایوبیز استفاده شده است که از روش TF-IDF[3] (به معنای فراوانی وزنی کلمه کلیدی) برای استخراج ویژگی بهره برده شده است [۷]. بخش دوم شامل آزمایش سه نوع یادگیری عمیق با اعمال جاسازی کلمه به عنوان بردار ورودی بود. نتایج به دست آمده از نظر دقت عملکرد بالاتر واضحی را برای رویکردهای یادگیری عمیق نشان داد.

الزیادی و همکاران دو مدل یادگیری عمیق شامل شبکه عصبی حافظه طولانی کوتاه مدت (LSTM)[4] و شبکه‌های عصبی کانولوشن (CNN)[5] پیاده‌سازی کردند [۸]. یک مدل ترکیبی از معماری CNN و شبکه عصبی بازگشتی (RNN)[6] ارائه شده

[1] Accuracy
[2] Precision
[3] Term Frequency -Inverse Document Frequency
[4] Long short-term memory networks
[5] Convolutional neural network
[6] Recurrent neural network

است که در این مدل ویژگی‌های محلی از طریق CNN به عنوان ورودی RNN برای تحلیل احساسات متون عربی جمع‌آوری شده است. طبق نتایج به دست آمده، مدل ترکیبی پیشنهادی (CNN-LSTM) عملکرد چشمگیری داشته است.

کیوان پور و همکاران یک روش مبتنی بر واژگان و یادگیری ماشین به نام OMLML[1] با استفاده از شبکه‌های اجتماعی پیشنهاد کردند [۹]. با توجه به روش پیشنهادی، ابتدا قطبیت نظرات نسبت به یک کلمه هدف با استفاده از روشی مبتنی بر واژگان و ویژگی‌های متنی کلمات و جملات تعیین شده است. در مرحله بعد، با نگاشت فضای ویژگی در یک بردار سه بعدی، نظرات بر اساس یک روش یادگیری ماشین جدید تجزیه و تحلیل و طبقه‌بندی شده‌اند. نتایج آزمایش‌های کمی و کیفی نشان داده است که نگاشت داده‌ها در فضای جدید هزینه آموزش را کاهش می‌دهد و عملکرد روش پیشنهادی به ویژه از نظر دقت و زمان اجرا قابل قبول بوده است.

در تحقیق سوزا و همکاران یک الگوریتم جدید برای عقیده کاوی و تجزیه و تحلیل احساسات متن ارسال شده در توییتر ارائه شده است که بر اساس خوشه‌بندی بدون نظارت عمل می‌کند. در این روش از نسخه ترکیبی بهینه‌سازی ازدحام ذرات (PSO) و جستجوی فاخته (CS) استفاده شده است [۱۰].

در مقاله داو و همکاران یک سیستم توصیه‌گر پیشنهاد شده است که از عقیده کاوی بر اساس رویکرد یادگیری عمیق برای بهبود دقت فرآیند توصیه استفاده کرده است [۱۱]. مدل پیشنهادی شامل دو قسمت می‌باشد. در قسمت ابتدایی، از یک شبکه عصبی پیچیده چند کانالی (MCNN)[2] برای استخراج بهتر جنبه‌ها و تولید رتبه‌بندی‌های جنبه خاص با محاسبه قطبیت احساسات کاربران در جنبه‌های مختلف استفاده شده است. در بخش دوم، رتبه‌بندی‌های جنبه خاص در یک ماشین تجزیه تانسور (TF) برای پیش‌بینی رتبه کلی ادغام شده است. نتایج تجربی با استفاده از مجموعه داده‌های مختلف نشان داده است که مدل پیشنهادی در مقایسه با روش‌های پایه به پیشرفت‌های قابل‌توجهی دست یافته است.

آخمدوا و همکاران از طبقه‌بندی‌کننده‌های مبتنی بر قانون فازی، شبکه‌های عصبی مصنوعی (ANN) و SVM برای عقیده کاوی استفاده کردند [۱۲]. برای تولید این روش‌ها، یک روش فراابتکاری اصلاح شده برای حل مسائل بهینه‌سازی متغیرهای واقعی یا باینری محدود و غیرمحدود پیشنهاد شده است. در این روش از طرح‌های وزن‌دهی اصطلاحات مختلف به عنوان رویکردهای پیش‌پردازش داده‌ها استفاده شده است.

کریشنا و همکاران مدل جدیدی برای عقیده کاوی و تحلیل احساسات با بررسی متن‌های ارسال شده در توییتر پیشنهاد کردند [۱۳]. مدل پیشنهادی در این مقاله از رویکردهای یادگیری ماشین و فازی برای عقیده‌کاوی و طبقه‌بندی احساسات در متن‌های مورد بررسی، استفاده کرده است.

در تحقیق نارایان و همکاران برای تشخیص و بازبینی هرزنامه، یک روش عقیده کاوی پیشنهاد شده است [۱۴]. در این روش از مجموعه‌های مختلفی از ویژگی‌ها، LIWC[3]، تگ‌های POS، ویژگی N-gram[4] و امتیاز احساسات استفاده شده است. برای طبقه‌بندی نظرات از شش روش درخت تصمیم، بیز ساده، SVM، K–نزدیکترین همسایه، جنگل تصادفی و رگرسیون لجستیک استفاده شده است.

کومار و همکاران بر روی عقیده کاوی وبگاه‌هایی مانند آمازون متمرکز شدند [۱۵]. وبگاه آمازون اجازه می‌دهد تا کاربر به صورت آزاد نظر خود را ارسال کند. سیستم پیشنهادی این تحقیق به طور خودکار، بررسی‌ها را از وبگاه استخراج می‌کند. پس از آن با استفاده از الگوریتم‌هایی مانند طبقه‌بندی نایوبیز، رگرسیون منطقی نظرات به عنوان مثبت و منفی طبقه‌بندی می‌کند. در پایان از متغیرهای متریک کیفیت برای اندازه‌گیری عملکرد هر الگوریتم استفاده شده است.

البلتاگی و همکاران یک مدل تجزیه و تحلیل احساسات مبتنی بر رویکردهای یادگیری ماشین را که ترکیبی از چندین ویژگی است، ارائه کردند [۱۶]. اکثر آنها با استفاده از احساسات واژگان

[1] Opinion mining method based on lexicon and machine learning
[2] Multi convolutional neural network
[3] Linguistic Inquiry and Word Count
[4] An N-Gram is a connected string of N items from a sample of text or speech

عربی استخراج شدند. مراحل مختلفی مانند تعیین طول متن، تعداد بخش‌ها در نظر گرفته شده است. افزایش دقت به دست آمده در شش مجموعه از هفت مجموعه داده مورد توجه قرار گرفته است. این سیستم بر روی مجموعه داده‌های رسانه‌های اجتماعی عربستان، مصر و شام اعمال شده و نتایج قابل قبولی به دست آمده است.

کریشنا و همکاران مدل جدیدی برای طبقه‌بندی و تحلیل احساسات پیشنهاد کردند [۱۷]. مدل پیشنهادی در این مقاله از رویکردهای یادگیری عمیق و رگرسیون بردار پشتیبان برای عقیده‌کاوی و طبقه‌بندی احساسات در متن‌های مورد بررسی استفاده کرده است. متغیرهای مورد استفاده برای ارزیابی، تعداد کلمات بررسی مشتری، دقت، پیچیدگی زمانی و نرخ مثبت کاذب بوده‌اند.

شایک و همکاران به تحلیل احساسات و نظرکاوی در مورد داده‌های آموزشی پرداختند [۱۸]. در این تحقیق (۱) نقش تحلیل عاطفی در آموزش را در چهار سطح در نظر گرفته است: سطح سند، سطح جمله، سطح نهاد و سطح جنبه، (۲) رویکردهای حاشیه‌نویسی احساسات شامل رویکردهای مبتنی بر واژگان و پیکره محور برای حاشیه‌نویسی‌های بدون نظارت بررسی گردیده، (۳) نقش هوش مصنوعی در تجزیه و تحلیل احساسات با روش‌هایی مانند یادگیری ماشین، یادگیری عمیق و تبدیل کننده‌ها مورد بحث قرار گرفته‌اند و (۴) تاثیر تجزیه و تحلیل احساسات بر رویه‌های آموزشی برای افزایش آموزش، ارزیابی و تصمیم‌گیری ارائه گردیده است.

علیان نژادی و همکاران یک روش جدید به منظور بازیابی تصویر مبتنی بر محتوا ارائه کردند [۱۹]. برای این منظور، با توجه به اهمیت بافت اشیا در یک تصویر، ویژگی جدیدی تحت عنوان هیستوگرام اختلاف بافت در جهت لبه برابر معرفی شده است. در روش پیشنهادی، ابتدا ویژگی‌هایی شامل ویژگی جدید معرفی شده، از تصاویر آموزشی استخراج شده و سپس تعدادی از این ویژگی‌ها انتخاب شده و با استفاده از آنها و کلاس هر تصویر و همچنین روش یادگیری ماشین بردار پشتیبان، تصاویر در کلاس‌های مختلف به سیستم آموزش داده شده‌اند.

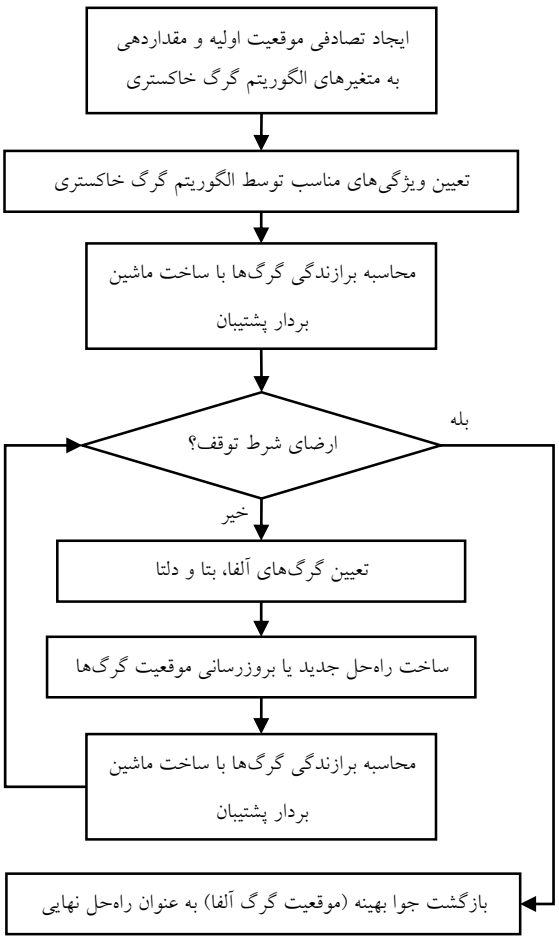
درخشی و همکاران دو مدل شبکه عصبی عمیق شامل شبکه عصبی پیچشی ParsCNN و شبکه عصبی با حافظه بلند کوتاه مدت دوسویه سلسه‌مراتبی با لایه توجه ParsBiLSTM برای طبقه‌بندی متون فارسی استفاده کردند [۲۰]. کارایی از نظر سه معیار ارزیابی دقت، فراخوانی و امتیاز F1 مورد مطالعه قرار گرفت. نتایج آزمایش‌ها نشان می‌دهد که روش ParsCNN میزان دقت ۰/۶۹، فراخوانی ۰/۷ همچنین روش ParsBiLSTM میزان دقت ۰/۷۲، فراخوانی ۰/۷۳ و امتیاز F1 برابر با ۰/۷۲ دارند که نشان‌دهنده کارایی بالاتر این روش‌ها نسبت به روش‌های طبقه‌بندی متون فارسی مورد مطالعه بوده است.

زارع و همکاران یک سیستم تجزیه و تحلیل احساسات بر روی داده‌های تلگرام فارسی پیشنهاد کردند [۲۱]. برای این منظور، چند روش استخراج ویژگی شامل بردار رخداد، فراوانی سند و ماتریس تعبیه کلمات جهت بازنمایی داده‌های متنی به عددی استفاده شده است و سپس طبقه‌بندی داده‌ها به روش‌های مختلف یادگیری ماشین کلاسیک شامل ماشین بردار پشتیبان، درخت تصمیم، بیز ساده و همچنین روش‌های یادگیری عمیق شامل شبکه عصبی عمیق، شبکه عصبی پیچشی و شبکه‌های حافظه طولانی کوتاه مدت یک‌طرفه و دوطرفه بررسی شده است. درنهایت ارزیابی و تحلیل نتایج نشان داده است که بهترین کارایی توسط روش استخراج ویژگی ماتریس تعبیه کلمات به همراه شبکه حافظه طولانی کوتاه مدت دوطرفه با دقت ۶۷/۹۰، صحت ۹۰/۰۱، فراخوان ۸۹/۵۴ و امتیاز F1 برابر با ۷۷/۸۹ درصد به دست آمده است.

۳. روش تحقیق

در این تحقیق یک رویکرد ترکیبی مبتنی بر روش‌های ماشین بردار پشتیبان و الگوریتم گرگ‌های خاکستری برای بهبود فرآیند عقیده کاوی ارائه شده است. بخش اصلی سیستم پیشنهادی، ماشین بردار پشتیبان است که الگوهای موجود در داده‌ها را با دریافت داده‌های آموزشی جستجو می‌کند. پس از پایان مرحله آموزش، سیستم می‌تواند برای عقیده کاوی بکار گرفته شود. برای افزایش کارایی روش پیشنهادی از الگوریتم گرگ‌های

خاکستری برای انتخاب مناسب‌ترین ویژگی‌ها استفاده می‌شود به طوری که نتایج به دست آمده مطلوب‌تر خواهد بود. نحوه ترکیب الگوریتم گرگ‌های خاکستری و ماشین بردار پشتیبان در شکل (۱) آورده شده است.



شکل (۱): نحوه ترکیب الگوریتم گرگ‌های خاکستری و ماشین بردار پشتیبان

تعیین ویژگی‌های موثر یکی از مهمترین مسائل در به دست آوردن کارایی مطلوب برای سیستم است. در این الگوریتم، هر فرد نشان‌دهنده تعدادی ویژگی است که در طول تکرارهای مختلف به سمت پاسخ بهینه حرکت می‌کند و در نهایت فردی انتخاب می‌شود که بهترین ویژگی‌ها را برای مساله انتخاب می‌کند. روشی که برای یادگیری تابع هدف الگوریتم بهینه‌سازی گرگ‌های خاکستری استفاده می‌شود، ماشین بردار پشتیبانی است که از ویژگی‌های مربوط به گرگ‌ها برای مدل‌سازی استفاده می‌کند. هدف الگوریتم گرگ‌های خاکستری پیدا کردن مناسب‌ترین ویژگی‌های ممکن می‌باشد، به طوری که ماشین بردار پشتیبان متناظر با آن بهترین عملکرد را داشته باشد. مطابق با شکل (۱) برای رسیدن به این هدف، در ابتدا موقعیت اولیه گرگ‌ها به صورت تصادفی ایجاد می‌شوند و به این ترتیب جمعیت اولیه ساخته و متغیرهای آن مقداردهی می‌شوند. برای محاسبه میزان تناسب هر گرگ در سیستم پیشنهادی، ماشین بردار پشتیبان با استفاده از ویژگی‌های انتخاب شده توسط گرگ ساخته می‌شود و دقت ماشین بردار پشتیبان تعیین می‌گردد. در مرحله بعد، گرگ‌های آلفا، بتا و دلتا ساخته می‌شوند. در ادامه، موقعیت گرگ‌ها با بروزرسانی موقعیت آنها تغییر می‌کند. پس از آن تناسب هر یک از گرگ‌ها دوباره مشخص خواهد شد. مراحل تعیین گرگ‌های آلفا، بتا و دلتا و بروزرسانی موقعیت گرگ‌ها تا زمانی که شرایط توقف الگوریتم خاتمه یابد، تکرار خواهد شد. در نهایت بهترین راه‌حل موجود به عنوان نتیجه نهایی ارائه خواهد شد و ماشین بردار پشتیبان برای آن ساخته شده و با داده‌های آموزش، آموزش می‌بیند. در نهایت دقت به دست آمده از این ماشین بردار پشتیبان به عنوان دقت سیستم پیشنهادی ارائه می‌شود.

۱.۳. ماشین بردار پشتیبان

ماشین بردار پشتیبان الگوریتمی برای طبقه‌بندی و دسته‌بندی کردن اشیا و داده‌ها می‌باشد و مبنای آن برای بهینه‌کردن، کم کردن خطای طبقه‌بندی داده‌های آموزشی می‌باشد [۲۲]. از ویژگی‌های ماشین‌های بردار پشتیبان می‌توان به موارد مانند طراحی دسته‌بندی‌کننده با بیشینه تصمیم، دستیابی به بهینه سراسری تابع هزینه، مشخص نمودن ساختار و توپولوژی دسته‌بندی‌کننده و پیدا کردن منطقه بهینه توسط مدل‌های غیرخطی اشاره کرد.

رویکرد ماشین بردار پشتیبان شکل ویژه‌ای از مدل‌های خطی را مورد استفاده قرار می‌دهد که یک منطقه حاشیه را حداکثر کنند. در واقع به دنبال حداکثر کردن، حداقل فاصله مرز تصمیم‌گیری از طبقه‌ها است. الگوریتم برای یافتن خط جداکننده طبقه‌ها از

دو خط موازی استفاده می‌کند که در خلاف جهت هم حرکت می‌کنند تا به یک طبقه برسند. در فاصله بین دو خط یک حاشیه به وجود می‌آید. هرچه حاشیه پهن‌تر باشد، یعنی هدف الگوریتم که حداکثر کردن حاشیه بوده است، محقق شده است. بیشینه کردن حاشیه ابر صفحه سبب حداکثر شدن جداسازی بین طبقه‌ها می‌گردد [۲۳]. ماشین بردار پشتیبانی یکی از روش‌های یادگیری بانظارت است که از آن برای طبقه‌بندی خطی و غیرخطی و نیز رگراسیون چندبعدی استفاده می‌کنند. اساس کاری کلاس‌بندی‌کننده ماشین بردار پشتیبان، کلاس‌بندی خطی داده‌ها است که در آن خطی انتخاب می‌شود که حاشیه اطمینان بیشتری داشته باشد. مساله پیدا کردن خط بهینه برای کلاس‌بندی داده‌ها به وسیله روش‌های برنامه‌ریزی غیرخطی که در حل مسائل محدودیت‌دار شناخته شده هستند، صورت می‌گیرد. برای اینکه ماشین بتواند داده‌های با پیچیدگی بالا را نیز دسته‌بندی کند، داده‌ها را به وسیله یک تبدیل کرنل به فضای با ابعاد خیلی بالاتر می‌برند.

۲.۳. الگوریتم گرگ خاکستری

الگوریتم گرگ‌های خاکستری با الهام از زندگی اجتماعی و شکار گرگ‌های خاکستری از چهار نوع گرگ برای شبیه‌سازی مراتب سلسله رهبری استفاده می‌کند. گرگ‌های خاکستری یک سلسله مراتب بسیار سخت اجتماعی دارند که شامل چهار رده آلفا، بتا، امگا و دلتا هستند. در اجتماع گرگ‌های خاکستری موقعیت هر گرگ مشخص است. آنها طعمه را محاصره می‌کنند و سپس با دستور رهبر (گرگ آلفا) به طعمه حمله می‌کنند. شکار شامل سه مرحله شناسایی، محاصره و حمله به طعمه است. گرگ آلفا رهبر گروه است که وظیفه تصمیم‌گیری در خصوص مسائل مختلف را به عهده دارد و سایر گرگ‌ها از دستورات وی تبعیت می‌کنند و به شکلی رفتار نیمه دموکراسی دارند. گرگ‌های آلفا تنها مجاز به جفت‌گیری در داخل گروه هستند. آلفا همیشه قوی‌ترین عضو گله نیست، ولی به لحاظ رهبری از سایر گرگ‌ها برتر است و به گرگ‌های ضعیف که توانایی شکار کردن ندارند نیز کمک می‌کند [۲۴].

رده دوم در سلسله مراتب گرگ‌ها، گرگ‌های بتا هستند. گرگ بتا، در تصمیم‌گیری‌ها به آلفا کمک کرده و از آن حمایت می‌کند و باعث ایجاد نظم در گروه می‌شود و بهترین گزینه برای جایگزینی آلفا است. پایین‌ترین سطح در سلسله مراتب گرگ‌ها مربوط به امگا است. آنها ضعیف‌ترین عضو گروه هستند که بیشتر نقش قربانی و پیش‌مرگ را دارند و وظیفه نگهداری از گرگ‌های جوان را دارند. آخرین افرادی در گروه هستند که اجازه خوردن طعمه را دارند. اگر گرگی آلفا، بتا و یا امگا نباشد، زیردست (و یا دلتا در برخی منابع) نامیده می‌شود. گرگ دلتا به لحاظ رتبه در سطح پایین‌تری نسبت به گرگ بتا قرار دارند، آنها گرگ‌های قوی هستند ولی قدرت رهبری ندارند و در سطح بالاتری نسبت به امگا قرار دارند. علاوه بر سلسله مراتب اجتماعی، شکار گرگ‌های خاکستری دارای سه مرحله ردیابی، تعقیب و نزدیک شدن به طعمه است. برای مدل‌سازی سلسله مراتب اجتماعی گرگ‌ها بهترین پاسخ را آلفا و از بین بهترین راه‌حل‌ها، دومین و سومین آنها را بتا و دلتا در نظر می‌گیریم. بقیه راه‌حل‌های کاندید را امگا در نظر می‌گیریم. بهینه‌سازی توسط آلفا، بتا و دلتا هدایت می‌شود و گروه چهارم از این سه گروه پیروی می‌کند. مدل‌سازی رفتار محاصره گرگ‌ها از روابط (۱) و (۲) استفاده می‌کند.

$$\vec{D} = \left|\vec{C} . \vec{X}_P(t) - \vec{X}(t)\right| \qquad (۱)$$

$$\vec{X}(t+1) = \vec{X}_P(t) - \vec{A}.\vec{D} \qquad (۲)$$

در این روابط، t تعداد دفعات تکرار، A و C بردارهای ضریب، $\vec{X}_P$ بردار موقعیت شکار و X بردار موقعیت است. برای محاسبه بردارهای A و C از روابط (۳) و (۴) استفاده می‌شود.

$$\vec{A} = 2\vec{a}.\overrightarrow{r_1} - \vec{a} \qquad (۳)$$

$$\vec{C} = 2\overrightarrow{r_2} \qquad (۴)$$

بردار a از ۲ تا ۰ به صورت خطی در طول دوره تکرار در هر دو فاز اکتشاف و بهره‌برداری کاهش می‌یابد. r_1 و r_2 بردارهای تصادفی مابین ۰ و ۱ هستند. با توجه به تصادفی بودن بردارهای r_1 و r_2، گرگ‌ها می‌توانند موقعیت خود را در داخل فضایی که

طعمه را در برگرفته، به صورت تصادفی تغییر دهند. روابط مربوط به آن در معادله (۵) و (۶) آمده است.

$$\begin{aligned}\overrightarrow{D_\alpha} &= \left|\overrightarrow{C_1}.\vec{X}_\alpha - \vec{X}\right| \\ \overrightarrow{D_\beta} &= \left|\overrightarrow{C_2}.\vec{X}_\beta - \vec{X}\right| \\ \overrightarrow{D_\delta} &= \left|\overrightarrow{C_3}.\vec{X}_\delta - \vec{X}\right|\end{aligned} \quad (۵)$$

$$\begin{aligned}\overrightarrow{X_1} &= \overrightarrow{X_\alpha} - \overrightarrow{A_1}.\overrightarrow{D_\alpha} \\ \overrightarrow{X_2} &= \overrightarrow{X_\beta} - \overrightarrow{A_2}.\overrightarrow{D_\beta} \\ \overrightarrow{X_3} &= \overrightarrow{X_\delta} - \overrightarrow{A_3}.\overrightarrow{D_\delta}\end{aligned} \quad (۶)$$

همین مفهوم را می‌توان به یک فضای جستجوی n بعدی تعمیم داد. در این حالت گرگ‌های خاکستری پیرامون بهترین راه‌حل به دست آمده در ابعادی بیشتر از ابعاد مکعب حرکت می‌کنند. همان‌طور که پیش‌تر گفته شد، شکار به وسیله آلفا رهبری می‌شود. برای مدل نمودن این رفتار سه مورد از بهترین راه‌های به حل دست آمده را ذخیره کرده و دیگر عوامل جستجو طبق رابطه (۷) وادار می‌شوند تا جایگاه خود را با در نظر گرفتن موقعیت بهترین عوامل جستجو تغییر دهند.

$$\overrightarrow{X(t+1)} = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3} \quad (۷)$$

در این الگوریتم پیاده‌سازی فاز بهره‌برداری یا حمله که در صورت متوقف شدن طعمه اتفاق می‌افتد، با کاهش مقدار متغیر a از ۲ به ۱ انجام می‌شود. مقدار A نیز وابسته به a است، از این رو کاهش پیدا می‌کند. با کاهش مقدار A گرگ‌ها مجبور به حمله به سمت طعمه می‌شوند. همچنین برای جلوگیری از گیر افتادن در حداقل محلی، فاز شناسایی ارائه می‌شود. گرگ‌ها در ابتدا برای شناسایی طعمه از هم دور می‌شوند و بعد از یافتن طعمه و برای حمله به آن به هم نزدیک شده و دور شکار حلقه می‌زنند. برای شبیه‌سازی این واگرایی از بردار A با مقادیر تصادفی بزرگتر از ۱ و یا کوچکتر از ۱- استفاده می‌شود.

جزء تاثیرگذار دیگر بر فرآیند شناسایی مقدار C است. مقدار این بردار عددی تصادفی در بازه ۰ و ۲ است. این مقدار تصادفی تاثیر موقعیت طعمه را در تعیین فاصله، شدت ($C > 1$) یا ضعف ($C < 1$) می‌بخشد [۲۵].

۳.۳. فرآیند ارزیابی در سیستم پیشنهادی

در این بخش، معیارهای مختلفی که برای ارزیابی الگوریتم‌های دسته‌بندی وجود دارند، مورد بحث قرار خواهند گرفت. برای سادگی، تعریف این معیارها برای یک مساله با دو دسته ارائه می‌شود. در ادامه، ابتدا نگاهی به مفاهیم مثبت واقعی و کاذب، منفی واقعی و کاذب خواهیم داشت. سپس مفهوم ماتریس درهم ریختگی[1] بیان می‌شود. پس از آن تعریف‌هایی از معیارهای مختلف ارزیابی مطرح خواهد شد.

مثبت کاذب[2] یا FP، پیش‌بینی‌هایی مرتبط با موارد نادرست است که به اشتباه به عنوان درست پیش‌بینی شده‌اند. نرخ مثبت کاذب به کسری از نمونه‌های منفی که به عنوان مثبت تشخیص داده شده‌اند، گفته می‌شود (رابطه (۸)).

$$False\ positive\ rate = \frac{FP}{FP + TN} \quad (۸)$$

منفی کاذب[3] یا FN پیش‌بینی‌هایی مرتبط با موارد درست است که به اشتباه به عنوان نادرست پیش‌بینی شده‌اند. نرخ منفی کاذب به کسری از نمونه‌های مثبت که به عنوان منفی تشخیص داده شده‌اند، گفته می‌شود (رابطه (۹)).

$$False\ negative\ rate = \frac{FN}{TP + FN} \quad (۹)$$

مثبت واقعی[4] یا TP به موارد مثبتی گفته می‌شود که به طور صحیح در کلاس مثبت تشخیص داده شده‌اند. نرخ مثبت واقعی به کسری از نمونه‌های مثبت گفته می‌شود که به درستی به عنوان مثبت طبقه‌بندی شده‌اند (رابطه (۱۰)).

$$True\ positive\ rate = \frac{TP}{TP + FN} \quad (۱۰)$$

منفی واقعی[5] یا TN به موارد منفی گفته می‌شود که به طور صحیح در کلاس منفی تشخیص داده شده‌اند. نرخ منفی واقعی به کسری از نمونه‌های منفی گفته می‌شود که به درستی به عنوان منفی طبقه‌بندی شده‌اند (رابطه (۱۱)) [۲۶].

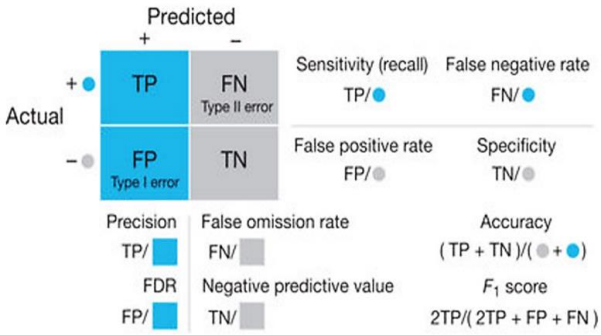
[1] Confusion matrix
[2] False Positive
[3] False Negative
[4] True Positive
[5] True Negative

$$True\ negative\ rate = \frac{TN}{TN + FP} \qquad (۱۱)$$

۴.۳. ماتریس پراکندگی

ماتریس پراکندگی، تعداد پیش‌بینی‌هایی درست و نادرست را در یک مجموعه داده شناخته شده نشان می‌دهد. شکل (۲) معرف یک ماتریس پراکندگی است و عناصر آن را شرح می‌دهد.

دایره‌های آبی و خاکستری به ترتیب نشان‌دهنده موارد واقعا مثبت $(TP + FN)$ و موارد واقعا منفی $(FP + TN)$ است. مربع‌های آبی و خاکستری نیز، به ترتیب موارد پیش‌بینی شده مثبت $(TP + FP)$ و پیش‌بینی شده منفی $(FN + TN)$ را نشان می‌دهد [۲۶].



شکل ۲ ماتریس پراکندگی و معیار های ارزیابی [۲۶]

۵.۳. معیارهای ارزیابی

از جمله مهمترین معیارهایی که برای ارزیابی الگوریتم‌های دسته‌بندی وجود دارند، معیارهای Accuracy، Precision، Recall و F1 هستند.

معیار دقت یا Accuracy مهمترین معیار برای تعیین کارایی یک الگوریتم دسته‌بندی می‌باشد و دقت کل دسته‌بندی را محاسبه می‌کند [۳]. این معیار با تقسیم تعداد مواردی که به عنوان مثبت شناسایی شده‌اند بر تعداد کل موارد مورد بررسی، به دست می‌آید (رابطه (۱۲)).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad (۱۲)$$

معیار Precision نشان‌دهنده نسبت کل موارد صحیح طبقه‌بندی شده در طبقه به کل موارد صحیح و غلط در طبقه می‌باشد (رابطه (۱۳)). این معیار بیان می‌کند که چند درصد از جواب‌های درست، درست بوده‌اند. به عبارتی، صحت دسته‌بندی دسته i را با توجه به کل موارد نشان می‌دهد. اندیس i در این معیار به این مفهوم است که معیار باید برای هر دسته i محاسبه شوند [۲۷].

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \qquad (۱۳)$$

معیار Recall نسبت موارد واقعی مثبت است که به طور صحیح مثبت پیش‌بینی شده است [۲۸]. در واقع فراخوانی معیاری است که مشخص می‌کند چه اندازه موارد صحیح درست تشخیص داده شده‌اند. حداقل مقدار آن صفر و حداکثر مقدار آن یک می‌باشد. نحوه محاسبه این معیار در رابطه (۱۴) آورده شده است.

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \qquad (۱۴)$$

معیار فراخوانی زمانی که ارزش منفی کاذب بالا باشد معیار خوبی است و کارایی دسته‌بندی را نشان می‌دهد و معیار صحت وقتی که ارزش مثبت کاذب بالا باشد معیار مناسبی است و دقت پیش‌بینی را نشان می‌دهد و اینکه مدل چقدر نتیجه را درست پیش‌بینی کرده است و تا چه اندازه می‌توان به خروجی اعتماد کرد [۲۹].

معیار F1، یک معیار مناسب برای ارزیابی دقت یک آزمایش است. این معیار Precision و Recall را با هم در نظر می‌گیرد (رابطه (۱۵)). معیار F1 در بهترین حالت، یک و در بدترین حالت صفر است.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (۱۵)$$

معیار Support یا پشتیبانی نشان‌دهنده اهمیت مجموعه داده بوده و شاخصی است که تعداد تراکنش‌های یک مجموعه آیتم یا اقلام را در پایگاه داده نشان می‌دهد و به نوعی یک محدودیت تکرار می‌باشد.

۴. یافته‌های پژوهش

۱.۴. معرفی مجموعه داده

برای پیاده‌سازی راه‌کار پیشنهادی، نیاز به مجموعه داده‌ها شامل

نظرات افراد می‌باشد. از این رو برای تهیه داده‌های مورد نیاز، از بسترهای برخط بهره برده شده است. یکی از این بسترها، وبگاه شناخته شده اسنپ تریپ[1] می‌باشد که در زیرمجموعه‌های خود، خدمات سفر و نقل و انتقالات را ارائه می‌دهد. به این ترتیب مجموعه نظرات وبگاه اسنپ تریپ به عنوان مجموعه داده مورد بررسی استفاده شده است. همچنین در این پژوهش از زبان برنامه‌نویسی پایتون و کتابخانه‌های هوش مصنوعی مطرح در این زبان بهره برده شده است.

۲.۴. آماده‌سازی داده

به دلیل آنکه متن‌ها اغلب شامل قالب‌های خاصی هستند که به متن کاوی کمک نمی‌کنند و می‌توانند حذف شوند، لازم است پیش از هرگونه تحلیل بر روی داده‌ها، بسته به نوع مساله قالب‌های غیرضروری حذف شوند. فرآیند پیش‌پردازش در این تحقیق از رویکردهایی چون حذف کلمات ایست واژه‌ها، حذف تک کاراکترها، حذف فاصله‌های اضافی، حذف علامت‌های نگارشی، اعمال TFIDF و ایجاد n-gram ها استفاده شده است. در فرآیند پیش‌پردازش متون، ابتدا ایست واژه‌ها[2] حذف می‌شوند. ایست واژه‌ها کلماتی هستند که به طور معمول به شکل گسترده در زبان مورد بررسی مورد استفاده قرار می‌گیرد و با وجود تکرار فراوان در متن از جهت معنایی اهمیت کمی دارند. این کلمات حاوی اطلاعات نیستند مانند ضمایر، حروف اضافه و حروف ربط. علاوه بر این، در مراحل بعد تک کارکترها و فاصله‌های اضافه موجود در متن‌ها و علامت‌های نگارشی حذف می‌شوند. پس از آن عملیات TFIDF برای تعیین میزان اهمیت کلمه‌ها نسبت به سند انجام می‌شود، در این مرحله تعداد تکرار کلمه در متن مورد بررسی قرار می‌گیرد هر چه یک کلمه بیشتر در یک متن تکرار شده باشد (TF) ولی در سایر متون کمتر تکرار شود (IDF)، مقدار TF-IDF آن بیشتر می‌شود و این معیار خوبی جهت تشخیص وزنِ یک کلمه در یک جمله است. در واقع نشان می‌دهد که یک کلمه چقدر می‌تواند یکتا و مهم باشد. سپس n-gram ها تعیین می‌شود در مدل n-gram برای ساخت کوله‌ای از کلمات (BoW)[3] علاوه بر اینکه هر کلمه، یک ستون ویژگی (یا بُعد) هست، ترکیب کلمات نیز به ابعاد اضافه می‌شوند. در نهایت با توجه به نامتوازن بودن مجموعه داده‌ها در این تحقیق، داده‌ها متوازن خواهند شد.

۳.۴. اعمال روش ماشین بردار پشتیبان

در این بخش، روش ماشین بردار پشتیبان بر روی مجموعه داده این تحقیق اعمال می‌شود. جدول (۱)، نتایج به دست آمده از این بخش را نشان می‌دهد. از اطلاعات موجود در این شکل می‌توان دریافت که با اعمال ماشین بردار پشتیبان بر روی داده‌ها، مقادیر Precision برای کلاس اول و دوم به ترتیب برابر با ۰/۶۰ و ۰/۸۶ می‌باشد. مقادیر Recall نیز برای کلاس اول و دوم به ترتیب برابر با ۰/۹۱ و ۰/۵۰ و مقدار Accuracy این روش برابر با ۰/۶۹ درصد می‌باشد.

جدول ۱: نتایج به دست آمده حاصل از پیاده سازی ماشین بردار پشتیبان

	Precision	Recall	F1-score	Support
Class 0	0.60	0.91	0.72	32
Class 1	0.86	0.50	0.63	38
Accuracy			0.69	70
macro avg	0.73	0.70	0.68	70
weighted avg	0.75	0.69	0.68	70

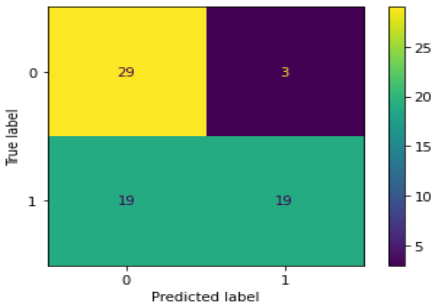
ماتریس درهم‌ریختگی به دست آمده حاصل از پیاده‌سازی ماشین بردار پشتیبان در شکل (۳) ارائه گردیده است. با توجه به نتایج حاصله از این ماتریس موارد به اشتباه تشخیص داده شده در دو کلاس به ترتیب برابر با ۱۹ و ۳ مورد هستند.

۴.۴. اعمال روش گرگ خاکستری و ماشین بردار پشتیبان

در این بخش از الگوریتم گرگ‌های خاکستری برای انتخاب ویژگی‌های موثر در مجموعه داده استفاده شده است. در طول اجرای الگوریتم گرگ‌های خاکستری هر راه‌حل نشان‌دهنده تعدادی ویژگی است. برای ارزیابی راه‌حل مورد نظر، ماشین بردار پشتیبان برای ویژگی‌های انتخاب شده پیاده‌سازی می‌شود.

[1] Snapptrip.com
[2] Stop Words
[3] Bag of Word

در نهایت بهترین راه‌حل نشان‌دهنده ویژگی‌هایی است که مناسب‌ترین کارایی را دارند. در این بخش، با بهره‌گیری از روش گرگ‌های خاکستری مناسب‌ترین ویژگی‌ها برای استفاده ماشین بردار پشتیبان تعیین می‌شود.



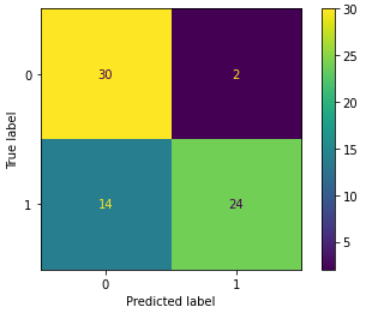
شکل (۳): ماتریس درهم‌ریختگی به دست آمده حاصل از پیاده‌سازی ماشین بردار پشتیبان

نتایج پیاده‌سازی روش ماشین بردار پشتیبان با گرگ خاکستری در جدول (۲) ارائه شده است. از نتایج حاصل می‌توان دریافت که با اعمال سیستم پیشنهادی بر روی داده‌ها، مقادیر Precision برای کلاس اول و دوم برابر با ۰/۶۸ و ۰/۹۲ می‌باشد. مقادیر Recall نیز برای کلاس اول و دوم به ترتیب برابر با ۰/۹۴ و ۰/۶۳ و مقدار Accuracy این روش برابر با ۰/۷۷ درصد است. مقایسه نتایج ثبت شده در این جدول نشان می‌دهد که روش پیشنهادی نسبت به طبقه‌بندی پایه به نتایج بهتری رسیده است.

جدول (۲): نتایج به دست آمده حاصل از پیاده‌سازی ماشین بردار پشتیبان و الگوریتم گرگ خاکستری

	Precision	Recall	F1-score	Support
Class 0	0.68	0.94	0.79	32
Class 1	0.92	0.63	0.75	38
Accuracy			0.77	70
macro avg	0.80	0.78	0.77	70
weighted avg	0.81	0.77	0.77	70

ماتریس درهم‌ریختگی به دست آمده حاصل از سیستم پیشنهادی برای عقیده کاوی در شکل (۴) ارائه گردیده است. با توجه به خروجی حاصله از این ماترس موارد به اشتباه تشخیص داده شده در هر دو کلاس تعداد کمتری را نسبت به قبل نشان می‌دهند. این مساله عملکرد مناسب سیستم پیشنهادی این تحقیق را در هر دو کلاس نشان می‌دهد.



شکل (۴) : ماتریس درهم‌ریختگی به دست آمده حاصل از پیاده‌سازی سیستم پیشنهادی

در ادامه نتایج حاصل از روش پیشنهادی به لحاظ دقت با سایر روش‌ها نظیر روش ترکیبی یادگیری عمیق و رگرسیون بردار پشتیبان (ICSD)[1] [۱۷]، روش ترکیبی شبکه عصبی حافظه طولانی کوتاه مدت (LSTM) و شبکه‌های عصبی کانولوشن (CNN) [۸]، روش ترکیبی بهینه‌سازی ازدحام ذرات (PSO)[2] و جستجوی فاخته (CS)[3] [۱۰]، روش ترکیبی الگوریتم فرااابتکاری COBRA[4] و الگوریتم SVM[5] [۱۲]، مقایسه گردید که نتایج در جدول (۳) قابل رویت می‌باشد.

جدول (۳): مقایسه نتایج روش پیشنهادی با سایر رویکردها

روش	دقت (درصد)
ICSD	78.8
LSTM-CNN	70.6
IDPSOMUT/CS	63.3
COBRA-SVM	62
Proposed Method	77

همان‌گونه که از جدول مقایسه نتایج روش پیشنهادی با سایر روش‌ها پیداست، دقت به دست آمده از روش پیشنهادی در قیاس با روش دوم (که در آن ویژگی‌های متن که از طریق CNN به دست آمده به عنوان ورودی برای شبکه‌های عصبی بازگشتی

[1] Lernt Independent Component Support Vector Regressive Deep
[5] Cuckoo Search
[6] Co-operation of Biology Related Algorithms
[7] Support Vector Machine

جهت تحلیل احساسات بکار برده شده) و روش سوم (که بر اساس یک الگوریتم PSO خودتطبیقی بهبودیافته با تابع تشخیص و جستجوی فاخته بوده است) و روش چهارم (که از روش‌های بهینه‌سازی محدود باینری و SVM استفاده نموده است)، دارای دقت بالاتری می‌باشد. در این مقاله، ما یک رویکرد موثر تحلیل عقیده کاوی را با استفاده از معماری مشترک ماشین بردار پشتیبان و گرگ خاکستری (GOW-SVM) پیشنهاد کردیم که در مقایسه با برخی روش‌های ترکیبی یادگیری ماشین عملکرد بهتری داشته است. در ادامه پیشنهاد می‌گردد با توجه به عملکرد مناسب مدل اول (یادگیری عمیق و رگرسیون بردار پشتیبان) که نسبت به سایر روش‌های بررسی شده سبب افزایش دقت و کاهش زمان برای استخراج کلمات گردیده است، می‌توان از الگوریتم گرگ خاکستری به دلیل کاهش ابعاد و افزایش سرعت همگرایی جهت تعیین ویژگی‌های مناسب متغیرها در ICSD استفاده نمود و بدین ترتیب بتوان به عملکرد بالاتری دست یافت.

۵. نتیجه‌گیری

امروزه اینترنت بشریت را تشویق می‌کند تا نظرات خود را از طریق وبلاگ‌ها، وبگاه‌ها و توییت‌ها منتشر کند. ارزیابی یک سرویس، یک محصول یا حتی یک فرد مشهور و ارسال نظر در مورد هر یک از آنها مفید است. از این رو، طبقه‌بندی نظر کاوی به یک حوزه تحقیقاتی جالب برای تجزیه و تحلیل خودکار این حجم عظیم از نظرات و وبلاگ‌ها تبدیل شده است. برای عقیده کاوی رویکردهای مختلفی توسط محققان در گذشته توسعه داده شده است. در این زمینه، رویکردهای مبتنی بر یادگیری ماشین به دلیل فرآیند یادگیری طبقه‌بندی‌کننده‌ها، امیدوارکننده‌ترین رویکرد در نظر گرفته شده‌اند. در این کار، روش ماشین بردار پشتیبان و رویکرد ترکیبی آن با الگوریتم بهینه‌سازی گرگ‌های خاکستری برای عقیده کاوی مورد بررسی قرار گرفت. نتایج پیاده‌سازی این سیستم‌ها نشان داد که تلفیق روش‌های مذکور می‌تواند نتایج مناسبی را به دنبال داشته باشد. در سیستم پیشنهادی این تحقیق، الگوریتم گرگ‌های خاکستری برای تعیین ویژگی‌های موثر در فرآیند عقیده کاوی بکار رفته است و عملکرد ماشین بردار پشتیبان را به نحو مطلوبی بهبود داد به گونه‌ای که عملکرد سیستم پیشنهادی در هر دو کلاس بهتر از گذشته شد.

تعارض منافع: نویسندگان اعلام می‌کنند که هیچ تعارض منافعی ندارند.

مراجع

[1] R.K. Bakshi, N. Kaur, R. Kaur, and G. Kaur, "Opinion mining and sentiment analysis," in 3rd Int. Conf. Comput. Sustain. Glob. Dev. (INDIACom), New Delhi, India, 2016, pp. 452-455.

[2] H. Chen and D. Zimbra, "AI and opinion mining," IEEE Intell. Syst., vol. 25, no. 3, pp. 74-80, 2010, doi: 10.1109/MIS.2010.75.

[3] S. Alimardani and A. Aghaie, "Opinion Mining in Persian Language," J. Inf. Technol. Manag., vol. 7, no. 2, pp. 345-362, 2015, doi: 10.22059/jitm.2015.53995 [In Persian].

[4] M.Y. Rau and. M.A. Kulkarni, "Polarity shift in opinion mining," in IEEE Int. Conf. Adv. Electronics Commun. Comput. Technol. (ICAECCT), Pune, India, 2016, pp. 333-337, doi: 10.1109/ICAECCT.2016.7942608.

[5] S.K. Rajan, A.F.S. Devaraj, R. Manickam, E.G. Julie, Y.H. Robinson, and V. Shanmuganathan, "Exploration of sentiment analysis and legitimate artistry for opinion mining," Multim. Tools Appl., vol. 81, no. 9, pp. 11989-12004, 2022, doi: 10.1007/s11042-020-10480-w.

[6] R. Ganesh, D. Saketh, V.D.V. Sai Kumar, M. Ramesh, "Website Evaluation Using Opinion Mining," in 3rd Int. Conf. Invent. Res. Comput. Appl. (ICIRCA), Coimbatore, India, 2021, pp. 1499-1506, doi: 10.1109/ICIRCA51532.2021.9544566.

[7] D.S.A. Elminaam, N. Neggaz, I.A.E. Gomaa, F.H. Ismail, and A.A. Elsawy, "ArabicDialects: An Efficient Framework for Arabic Dialects Opinion Mining on Twitter Using Optimized Deep Neural Networks," IEEE Access, vol. 9, pp. 97079-97099, 2021, doi: 10.1109/ACCESS.2021.3094173.

[8] H. Elzayady, K.M. Badran, and G.I. Salama, "Arabic Opinion Mining Using Combined CNN-LSTM Models," Int. J. Intell. Syst. Appl., vol. 12, no. 4, 2020,

doi: 10.5815/ijisa.2020.04.03.

[9] M. Keyvanpour, Z.K. Zandian, and M. Heidarypanah, "OMLML: a helpful opinion mining method based on lexicon and machine learning in social networks," Soc. Netw. Anal. Min., vol. 10, no. 1, p. 10, 2020, doi: 10.1007/s13278-019-0622-6.

[10] E. Souza, D Santos, G.H.F.M. Oliveira, A. Silva, and A.L.I. Oliveira, "Swarm optimization clustering methods for opinion mining," Nat. Comput., vol. 19, no. 3, pp.547-575, 2020, doi: 10.1007/s11047-018-9681-2.

[11] A. Da'u, N. Salim, I. Rabiu, and A. Osman, "Recommendation system exploiting aspect-based opinion mining with deep learning method," Inf. Sci., vol. 512, pp. 1279-1292, 2020, doi: 10.1016/j.ins.2019.10.038.

[12] S. Akhmedova, E. Semenkin, and V. Stanovov, "Co-operation of biology related algorithms for solving opinion mining problems by using different term weighting schemes," in 13th Int. Conf. Inf. Control Autom. Robot. (ICINCO), Lisbon, Portugal, 2016, pp. 73-90, doi: 10.1007/978-3-319-55011-4_4.

[13] B.V. Krishna, A.K. Pandey, and A.P.S. Kumar, "Feature based opinion mining and sentiment analysis using fuzzy logic," in Cognitive Science and Artifcial Intelligence, Singapore, pp. 79-89, Springer, 2018, doi: 10.1007/978-981-10-6698-6_8.

[14] R. Narayan, J.K. Rout, and S.K. Jena, "Review spam detection using opinion mining," in Prog. Intel. Comput. Techn. Theory Pract. Appl., Springer, Singapore, 2018, pp. 273-279, doi: 10.1007/978-981-10-3376-6_30.

[15] K.L.S. Kumar, J. Desai, and J. Majumdar, "Opinion mining and sentiment analysis on online customer review," in IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC), Chennai, India, 2016, pp. 1-4, doi: 10.1109/ICCIC.2016.7919584.

[16] S.R. El-Beltagy, T. Khalil, A. Halaby, and M. Hammad, "Combining lexical features and a supervised learning approach for Arabic sentiment analysis," in Comput. Linguist. Intell. Text Process. 17th Int. Conf. (CICLing), Konya, Turkey, 2018, Part II 17, pp. 307-319, doi: 10.1007/978-3-319-75487-1_24.

[17] M.M. Krishna, B. Duraisamy, and J. Vankara, "Independent component support vector regressive deep learning for sentiment classification," Meas. Sensors, vol. 26, p. 100678, 2023, doi: 10.1016/j.measen.2023.100678.

[18] M.M. AlyanNezhadi, M. Hosseini, H. Qazanfari, and A. Kamandi, "Content-based image retrieval using support vector machine and texture difference histogram features," Soft Comput. J., vol. 11, no. 1, pp. 10-21, 2022, doi: 10.22052/scj.2022.246175.1053 [In Persian].

[19] M. Feizi-Derakhshi, Z. Mottaghinia, and M. Asgari-Chenaghlu, "Persian Text Classification Based on Deep Neural Networks," Soft Comput. J., vol. 11, no. 1, pp. 120-139, 2022, doi: 10.22052/scj.2023.243182.1010 [In Persian].

[20] F. Zare Mehrjardi, M. Yazdian-Dehkordi, and A. Latif, "Evaluating classical machine learning and deep-learning methods in sentiment analysis of Persian telegram message," Soft Comput. J., vol. 11, no. 1, pp. 88-105, 2022, doi: 10.22052/scj.2023.246553.1077 [In Persian].

[21] T. Shaik, X. Tao, C. Dann, H. Xie., Y. Li, and L. Galligan, "Sentiment analysis and opinion mining on educational data: A survey," Nat. Lang. Process. J., vol. 2, p. 100003, 2023, doi: 10.1016/j.nlp.2022.100003.

[22] R. Raei and S. Fallahpour, "Support Vector Machines Application in Financial Distress Prediction of Companies Using Financial Ratios," Account. Audit. Rev., vol. 15, no. 4, pp. 17-34, 2009, dor: 20.1001.1.26458020.1387.15.4.2.6 [In Persian].

[23] K.-S. Shin, T.S. Lee, and H.-J. Kim, "An application of support vector machines in bankruptcy prediction model," Expert Syst. Appl., vol. 28, no. 1, pp. 127-135, 2005, doi: 10.1016/j.eswa.2004.08.009.

[24] S. Mirjalili, S.M. Mirjalili, and A. Lewis, "Grey wolf optimizer," Adv. Eng. Softw., vol. 69, pp. 46-61, 2014, doi: 10.1016/j.advengsoft.2013.12.007.

[25] A. Mohammadzadeh, M. Masdari, F.S. Gharehchopogh, and A. Jafarian, "An improved grey wolves optimization algorithm for workflow scheduling in cloud computing environment," J. Soft Comput. Inf. Technol., vol. 8, no. 4, pp. 17-29, 2019 [In Persian].

[26] J. Lever, M. Krzywinski, and N. Altman, "Classification evaluation," Nat. Methods, vol. 13, pp. 603-604, 2016, doi: 10.1038/nmeth.3945.

[27] G. Xu, Y. Zong, and Z. Yang, Applied data mining, CRC Press, 2013.

[28] D.M.W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," arXiv, CoRR abs/2010.16061, 2020, doi: 10.48550/arXiv.2010.16061.

[29] M. Khodayar and A. Osare, "Intrusion detection in computer networks using hybrid machine learning techniques,"in 9th Int. Symp. Adv. Sci. Technol., Mashhad, Iran, 2014 [In Persian].