بهبود فرآیند عقیده‌کاوی به کمک تلفیق روش های گرگ خاکستری و ماشین بردار پشتیبان

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 پلتفرم‌های 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 تکنیک‌های مختلف هوش مصنوعی برای وظایف و برنامه‌های مختلف مرتبط با عقیده کاوی اعمال می‌شوند. در این مقاله رویکرد جدیدی مبتنی بر تکنیک ماشین بردار پشتیبان و الگوریتم بهینه‌سازی گرگ های خاکستری برای بهبود فرآیند عقیده کاوی پیشنهاد شده است، بگونه‌ای که الگوریتم گرگ های خاکستری برای تعیین ویژگی‌های موثر در فرآیند عقیده کاوی به کار رفته است و عملکرد ماشین بردار پشتیبان را بهبود می‌دهد. نتایج پیاده سازی این تحقیق نشان داد که سیستم پیشنهادی توانسته با انتخاب ویژگی‌های موثر به افزایش دقت و پوشاندن خطای تکنیک 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.

واژه‌های کلیدی: عقیده کاوی، ماشین بردار پشتیبان، الگوریتم بهینه سازی گرگ های خاکستری.

Improve opinion mining by combining gray wolf and support vector machine

Abstract: *The advent of the web and its continuous growth has created a tremendous amount of user-generated information.Valuable subjective information is easy to find. Especially on social networks and e-commerce platforms that contain essential information. As a result, the field of belief mining has attracted considerable attention in recent years. New research papers are published every day in which various AI techniques are applied to various tasks and applications related to mining opinion. In this article, a new approach based on the support vector machine technique and the gray wolf optimization algorithm is proposed to improve the opinion mining process. Such that, the gray wolf algorithm is used to determine the practical features in the belief mining process and enhances the performance of the support vector machine. This research showed that the proposed system could increase the accuracy and cover the error of the backup vector technique by selecting practical features. The proposed approach has been evaluated using three criteria: accuracy, recall, and precision. Precision for the first and second classes is 0.68 and 0.92%, respectively, and recall for the first and second classes equals 0.94 and 0.63%, and accuracy was 0.77%. The results indicate that the proposed system of this research achieved favorable results in both categories.*

Keywords: Opinion Mining, Support Vector Machine, Gray Wolf Optimization Algorithm.

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 فزاینده 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 سروکار دارد، حداقل تا حدی به عنوان پاسخی مستقیم به افزایش علاقه به سیستم‌های جدید رخ داده است که مستقیماً با نظرات به عنوان یک شیء درجه یک سروکار دارند[۱].

نظرکاوی، زیرشاخه‌ای در داده‌کاوی و زبان شناسی محاسباتی است که به تکنیک‌های محاسباتی برای استخراج، طبقه‌بندی، درک و ارزیابی نظرات بیان شده در منابع مختلف خبری آنلاین، نظرات رسانه‌های اجتماعی و سایر محتوای تولید شده توسط کاربر اشاره دارد. تحلیل احساسات اغلب در نظر کاوی برای شناسایی احساسات، عاطفه، ذهنیت و سایر حالات احساسی در متن آنلاین استفاده می‌شود[۲]. عقیده‌ها که بصورت داده‌های متنی جمع‌آوری می‌شوند، دارای اطلاعات ارزشمندی هستند که با توجه زمینه کاربرد، می‌توانند اطلاعات مفیدی را برای افراد و سازمان‌ها تهیه کنند. عقیده‌کاوی بدنبال مشخص نمودن تعیین جهت عقیده‌ها یا به عبارتی دیگر، مثبت و منفی بودن هر عقیده است[۳]. طبقه‌بندی نظرات یا احساسات، مهم‌ترین وظیفه در حوزه ی عقیده‌کاوی است، با توجه به اینکه در سال‌های اخیر، ثبت نظرات مردمی به صورت آنلاین به طور چشمگیری افزایش یافته است. به دست آوردن نتایج عقیده کاوی یکی زمینه‌های اصلی محاسبات احساسات می‌باشد [۴].

الگوریتم‌های متعددی در زمینهٔ عقیده کاوی توسعه یافته که هر یک در بهبود مسئله‌ی عقیده‌کاوی تلاش کرده‌اند. در این مقاله نیز هدف آن است که یکی از جدیدترین روش‌های فراابتکاری ارائه شده را به منظور عقیده کاوی به کار بگیریم. علت انتخاب این روش طبقه‌بندی، سادگی به کارگیری آن در حل مسائل درک نظرات می‌باشد. رویکرد جدید با ترکیب ماشین بردار پشتیبان و الگوریتم گرگ خاکستری ارائه گردیده که به دنبال عقیده کاوی و تجزیه و تحلیل احساسات موجود در متن‌ها می‌باشد و عقیده ی مثبت و منفی موجود در متن‌ها را مشخص می‌کند. انتخاب ویژگی با الگوریتم گرگ خاکستری سبب کاهش ابعاد و همچنین 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 عقیده کاوی در سطح وب و رسانه‌های اجتماعی به شکل‌دهی مجدد کسب‌وکارها و تحت تأثیر قرار دادن احساسات و عواطف عمومی کمک کرده است و عمیقاً بر سیستم‌های اجتماعی و سیاسی ما تأثیر گذاشته است. بنابراین جمع آوری و مطالعه‌ی 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. مجموعه ویژگی‌ها، با کمک رویکرد مبتنی بر معنای معرفی شده و به منظور در نظر گرفتن توانایی پیش‌بینی فردی کلمات و به حداقل رساندن ویژگی-های انتخاب شده مورد بررسی قرار گرفتند[۵]. گانش و همکاران از انواع طبقه‌بندی‌کننده‌ها مانند نایوبیز، جنگل تصادفی، درخت تصمیم، ماشین بردار پشتیبان، K نزدیکترین همسایه، رگرسیون لجستیک برای طبقه‌بندی متن استفاده کردند. در این مقاله طبقه‌بندی‌کننده رگرسیون لجستیک برای طبقه‌بندی متن و ایجاد حس کلمه بالاترین صحت[۱] ، یادآوری و امتیازF1 را در میان طبقه‌بندی‌کننده‌های دیگر به دست آورد. امتیاز F1 میانگین وزنی دقت و یادآوری است. امتیاز F1 معمولاً مفیدتر از دقت است، به خصوص اگر توزیع کلاس ناهموار باشد. علاوه بر این طبقه‌بندی کننده نایوبیز بهترین دقت[۲] را در بین طبقه‌بندی‌کننده‌های مورد

[۱] Accuracy
[۲] precision

استفاده به دست آورد[۶]. در مقاله آبدلمینام و همکاران، در بخش اول، از شش روش یادگیری ماشین از جمله درخت‌های تصمیم، رگرسیون لجستیک، نزدیک‌ترین همسایه، جنگل‌های تصادفی، ماشین‌های بردار پشتیبان و نایوبیز استفاده شد که از روش TF-IDF[۱](به معنای فراوانی وزنی کلمه کلیدی) برای استخراج ویژگی استفاده شد. بخش دوم شامل آزمایش سه نوع یادگیری عمیق با اعمال جاسازی کلمه به عنوان بردار ورودی بود. نتایج به‌دست‌آمده از نظر دقت عملکرد بالاتر واضحی را برای تکنیک‌های یادگیری عمیق نشان داد[۷]. الزیادی و همکاران دو مدل یادگیری عمیق شامل شبکه‌ی‌عصبی حافظه‌ی طولانی کوتاه‌مدت (LSTM[۲]) و شبکه‌های عصبی کانولوشن (CNN[۳]) پیاده‌سازی کردند. یک مدل ترکیبی از معماری CNN و شبکه عصبی بازگشتی (RNN[۴]) ارائه شد که در این مدل ویژگی‌های محلی از طریق CNN به عنوان ورودی RNN برای تحلیل احساسات متون عربی جمع‌آوری شد. طبق نتایج به دست آمده مدل ترکیبی پیشنهادی (CNN-LSTM) عملکرد چشمگیری داشته است.[۸].

کیوان پور و همکاران یک روش مبتنی بر واژگان و یادگیری ماشین به نام OMLML[۵] با استفاده از شبکه‌های اجتماعی پیشنهاد کردند. 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 شد. در مرحله بعد، با نگاشت فضای ویژگی در یک بردار ۳ بعدی، نظرات براساس یک روش یادگیری ماشین جدید تجزیه و تحلیل و طبقه‌بندی شدند. نتایج آزمایش‌های کمی و کیفی نشان داد که نگاشت داده‌ها در فضای جدید هزینه آموزش را کاهش می‌دهد و عملکرد روش پیشنهادی به‌ویژه از نظر دقت و زمان اجرا قابل قبول بوده است[۹]. در تحقیق سوزا و همکاران یک الگوریتم جدید برای عقیده کاوی و تجزیه و تحلیل احساسات متن ارسال شده در توییتر ارائه شد که بر اساس خوشه-بندی بدون نظارت عمل می‌کرد. در این روش از نسخه ترکیبی بهینه‌سازی ازدحام ذرات (PSO) و جستجوی فاخته (CS) استفاده شده بود[۱۰]. در مقاله داو و همکاران یک سیستم توصیه‌گر پیشنهاد شده است که از عقیده کاوی بر اساس تکنیک یادگیری عمیق برای بهبود دقت فرآیند توصیه استفاده کرده است. مدل پیشنهادی شامل دو قسمت می‌باشد. در قسمت ابتدایی، از یک شبکه عصبی پیچیده چند کانالی (MCNN)[۶] برای استخراج بهتر جنبه‌ها و تولید رتبه‌بندی‌های جنبه خاص با محاسبه قطبیت احساسات کاربران در جنبه‌های مختلف استفاده شد. در بخش دوم، رتبه‌بندی‌های جنبه خاص در یک ماشین تجزیه تانسور (TF) برای پیش‌بینی رتبه کلی ادغام شد. نتایج تجربی با استفاده از مجموعه داده‌های مختلف نشان داد که مدل پیشنهادی در مقایسه با روش‌های پایه به پیشرفت‌های قابل‌توجهی دست یافته است[۱۱]. آخمدوا و همکاران از طبقه‌بندی‌کننده‌های مبتنی بر قانون فازی، شبکه‌های عصبی مصنوعی (ANN) و SVM برای عقیده کاوی استفاده کردند. برای تولید این روش‌ها، یک روش فراابتکاری اصلاح‌شده برای حل مسائل بهینه‌سازی پارامترهای واقعی یا باینری محدود و غیرمحدود پیشنهاد شد. در این روش از طرح‌های وزن-دهی اصطلاحات مختلف به عنوان تکنیک های پیش پردازش داده-ها استفاده شد.[۱۲].

کریشنا و همکاران مدل جدیدی برای عقیده کاوی و تحلیل احساسات با بررسی متن‌های ارسال شده در توییتر پیشنهاد کردند. مدل پیشنهادی در این مقاله از تکنیک‌های یادگیری ماشین و رویکرد فازی برای عقیده‌کاوی و طبقه‌بندی احساسات در متن‌های مورد بررسی استفاده کرده است. [۱۳]. در تحقیق نارایان و همکاران برای تشخیص و بازبینی هرزنامه، یک روش عقیده کاوی پیشنهاد شده است. در این روش از مجموعه‌های مختلفی از ویژگی‌ها، LIWC[۷]، تگ‌های POS، ویژگی N-gram[۸] و امتیاز احساسات استفاده شده است. برای طبقه بندی نظرات از شش

[۱] Term Frequency -Inverse Document Frequency

[۲] long short-term memory networks

[۳] convolutional neural network

[۴] recurrent neural network

[۵] opinion mining method based on lexicon and machine learning

[۶] Multi convolutional neural network

[۷] Linguistic Inquiry and Word Count

[۸] An N-Gram is a connected string of N items from a sample of text or speech

تکنیک درخت تصمیم، بیز ساده، SVM، k-نزدیکترین همسایه (KNN)، جنگل تصادفی و رگرسیون لجستیک استفاده شده است[۱۴]. کومار و همکاران بر روی عقیده کاوی وب سایت‌هایی مانند آمازون متمرکز شدند. وب سایت آمازون اجازه می‌دهد تا کاربر به صورت آزاد نظر خود را ارسال کند. سیستم پیشنهادی این تحقیق به طور خودکار، بررسی‌ها را از وب سایت استخراج می-کند. پس از آن با استفاده از الگوریتم‌هایی مانند طبقه بندی نایو بیز، رگرسیون منطقی نظرات به عنوان مثبت و منفی طبقه بندی می‌کند. در پایان از پارامترهای متریک کیفیت برای اندازه گیری عملکرد هر الگوریتم استفاده شده است[۱۵]. البلتاگی و همکاران یک مدل تجزیه و تحلیل احساسات مبتنی بر تکنیک‌های یادگیری ماشین را که ترکیبی از چندین ویژگی است، ارائه کردند. اکثر آن‌ها با استفاده از احساسات واژگان عربی استخراج شدند. مراحل مختلفی مانند تعیین طول متن، تعداد بخش‌ها در نظر گرفته شد. افزایش دقت به دست آمده در شش مجموعه از هفت مجموعه داده مورد توجه قرار گرفت. این سیستم بر روی مجموعه داده‌های رسانه‌های اجتماعی عربستان، مصر، شام اعمال شد و نتایج قابل قبولی به دست آمد[۱۶].

کریشنا و همکاران مدل جدیدی برای طبقه‌بندی و تحلیل احساسات پیشنهاد کردند. مدل پیشنهادی در این مقاله از تکنیک‌های یادگیری عمیق و رگرسیون بردار پشتیبان برای عقیده‌کاوی و طبقه‌بندی احساسات در متن‌های مورد بررسی استفاده کرده است، پارامترهای مورد استفاده برای ارزیابی، تعداد کلمات بررسی مشتری، دقت، پیچیدگی زمانی و نرخ مثبت کاذب بوده‌اند[۱۷]. شایک و همکاران به تحلیل احساسات و نظرکاوی در مورد داده‌های آموزشی پرداختند. 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) تکنیک‌های حاشیه نویسی احساسات شامل رویکردهای مبتنی بر واژگان و پیکره محور برای حاشیه نویسی‌های بدون نظارت بررسی گردیده (۳) نقش هوش مصنوعی در تجزیه و تحلیل احساسات با روش‌هایی مانند یادگیری ماشین، یادگیری عمیق، و تبدیل کننده ها مورد بحث قرار گرفته اند (۴) تاثیر تجزیه و تحلیل احساسات بر رویه-های آموزشی برای افزایش آموزش، تصمیم‌گیری و ارزیابی ارائه گردیده است[۱۸].علیان نژادی و همکاران یک روش جدید به منظور بازیابی تصویر مبتنی بر محتوا کردند. برای این منظور، با توجه به اهمیت بافت اشیاء 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 این ویژگی‌ها و کلاس هر تصویر و همچنین روش یادگیری ماشین بردار پشتیبان، تصاویر در کلاس‌های مختلف به سیستم آموزش داده شده اند[۱۹]. 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. کارایی از نظر سه معیار ارزیابی دقت، فراخوانی و مقیاس F-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 مقیاس F 72/0-دارند که نشان‌دهنده کارایی بالاتر این روش‌ها نسبت به روش‌های طبقه‌بندی متون فارسی مورد مطالعه بوده است[۲۰]. زارع و همکاران یک سیستم تجزیه و تحلیل احساسات بر روی داده‌های تلگرام فارسی پیشنهاد کردند. برای این منظور، چند روش استخراج ویژگی شامل بردار رخداد، فراوانی سند و ماتریس تعبیه کلمات جهت بازنمایی داده‌های متنی به عددی گردید و سپس طبقه‌بندی داده‌ها 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 شد. در نهایت ارزیابی و تحلیل نتایج نشان داد 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 معیار F ،۷۷/۸۹ درصد به دست آمده است[۲۱].

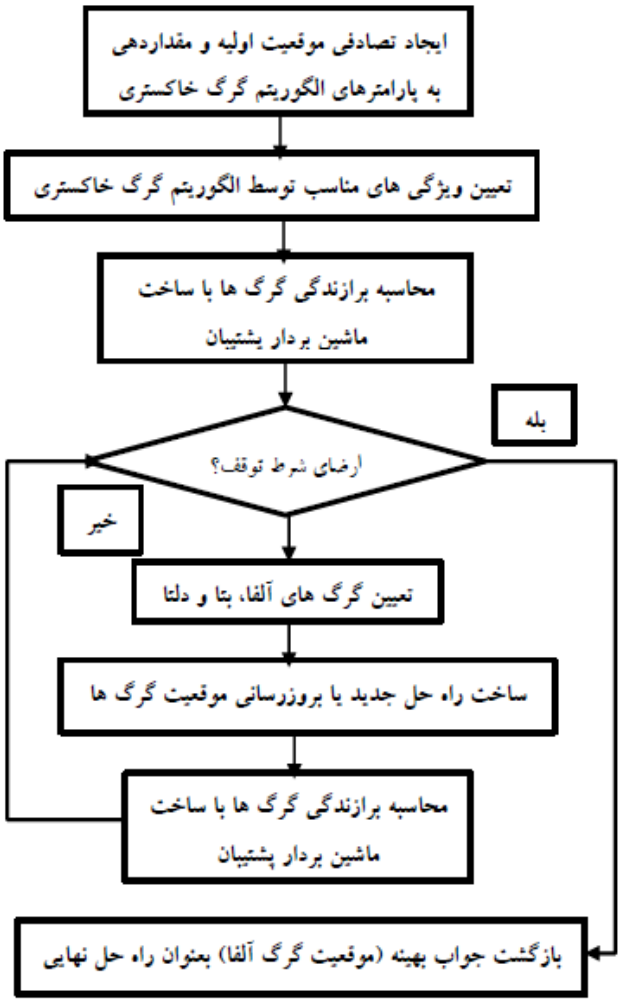
۳. روش تحقیق

در این تحقیق یک روش 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 مرحله‌ی آموزش، سیستم می‌تواند برای عقیده کاوی به کار گرفته شود. برای افزایش کارایی روش پیشنهادی از الگوریتم گرگ های خاکستری برای انتخاب مناسب‌ترین ویژگی‌ها استفاده می‌شود به طوری که نتایج به دست آمده مطلوب‌تر خواهد بود. نحوه‌ی ترکیب الگوریتم گرگ های خاکستری و ماشین بردار پشتیبان در شکل شماره ۱ آورده شده است.

تعین 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 مسئله انتخاب می‌کند. روشی که برای یادگیری تابع هدف الگوریتم بهینه‌سازی گرگ‌های خاکستری استفاده می‌شود، ماشین بردار پشتیبانی است که از ویژگی‌های مربوط به آن گرگ 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 اولیه‌ی گرگ‌ها به صورت تصادفی ایجاد می‌شوند و به این ترتیب جمعیت اولیه ساخته و پارامترهای آن مقداردهی می‌شود. برای محاسبه‌ی 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 مرحله‌ی بعد، گرگ‌های آلفا، بتا و دلتا ساخته می‌شوند. در مرحله‌ی بعد موقعیت گرگ‌ها با به روز رسانی موقعیت آن‌ها تغییر می‌کند. پس از آن تناسب هر یک از گرگ‌ها دوباره مشخص خواهد شد. مراحل تعیین گرگ‌های آلفا، بتا و دلتا و به روز رسانی موقعیت گرگ‌ها تا زمانی که شرایط توقف الگوریتم خاتمه یابد، تکرار خواهد شد. در نهایت بهترین راه حل موجود به عنوان نتیجه‌ی نهایی ارائه خواهد شد و ماشین بردار پشتیبان برای آن ساخته شده و با داده‌های آموزش، آموزش می‌بینند. در نهایت دقت به دست آمده از این ماشین بردار پشتیبان به عنوان دقت سیستم پیشنهادی ارائه می‌شود.



شکل ۱: نحوه‌ی ترکیب الگوریتم گرگ‌های خاکستری و ماشین بردار پشتیبان

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 خطـای طبقـه‌بنـدی داده‌هـای آموزشـی می‌باشد[۲۲]. از ویژگی های ماشین های بردار پشتیبان می‌توان به موارد زیر اشاره کرد:

طراحی دسته‌بندی کننده با ماکسیمم تصمیم، دستیابی به بهینه سراسری تابع هزینه، مشخص نمودن ساختار و توپولوژی دسته-بندی کننده، پیدا کردن منطقه بهینه توسط مدل‌های غیرخطی .

رویکرد ماشین بردار پشتیبان شکل ویژه ای 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 بوجود می‌آید هرچه حاشیه پهن‌تر باشد، یعنی هدف الگوریتم که حداکثر کردن حاشیه بوده است محقق شده است. بیشینه کـردن حاشـیه ابـر صـفحه سبب حـداکثر شـدن 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 کلاس‌بندی کنندهٔ ماشین بردار پشتیبان کلاس‌بندی خطی داده‌ها است و در کلاس‌بندی خطی داده‌ها 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 شناخته‌شده 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 گرگ های خاکستری موقعیت هر گرگ مشخص است انها طعمه را محاصره می‌کنند و سپس با دستور رهبر(آلفا) به طعمه حمله می‌کنند.شکار شامل سه مرحله شناسایی، محاصره و حمله به طعمه است. جفت آلفا رهبران گروه هستند که وظیفه تصمیم‌گیری در خصوص مسائل مختلف را به عهده دارند و سایر گرگ‌ها از دستورات وی تبعیت می‌کنند. و به شکلی رفتار نیمه دموکراسی دارند. گرگ‌های آلفا تنها مجاز جفت-گیری در داخل گروه هستند. آلفا همیشه قوی‌ترین عضو گله نیست، ولی به لحاظ رهبری از سایر گرگ‌ها برتر است و به گرگ-های ضعیف که توانایی شکار کردن ندارند نیز کمک می‌کند [۲۴].

رده دوم در سلسله مراتب یه دسته، بتا است. گرگ بتا، در تصمیم‌گیری‌ها به آآلفا کمک کرده و از آن حمایت می‌کند باعث ایجاد نظم و انظباط در گروه می‌شود و بهترین گزینه برای جایگزینی آلفا است. پایین ترین سطح در سلسله مراتب گرگ‌ها مربوط به امگا است. آنها ضعیف‌ترین عضو گروه هستند که بیشتر نقش قربانی و پیش مرگ را دارند و وظیفه نگهداری از گرگ‌های جوان را دارند. آخرین افرادی در گروه هستند که اجازه خوردن طعمه را دارند. اگرگرگی آلفا، بتا و یا امگا نباشد، زیردست (و یا دلتا در برخی منابع) نامیده می‌شود. گرگ دلتا به لحاظ رتبه در سطح پایین تری نسبت به بتا قراردارند، آنها گرگ‌های قوی هستند ولی قدرت رهبری ندارندو 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 راه حل‌ها، دومین و سومین را بتا و دلتا در نظر می‌گیریم. بقیه راه حل‌های کاندیدا را امگا در نظر می‌گیریم. بهینه-سازی 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 ۳و ۴ استفاده می‌شود.

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \quad (۳)$$

$$\vec{C} = 2\vec{r}. \quad (۴)$$

بردار a از ۲ تا ۰ به صورت خطی در طول دوره تکرار در هر دو فاز اکتشاف و بهره برداری کاهش می یابد. rبردار تصادفی مابین ۰ و ۱ است. با توجه به تصادفی بودن بردارهای r1 وr2 ،گرگ‌ها می‌توانند موقعیت خود را در داخل فضایی که طعمه را در برگرفته به صورت تصادفی تغییر دهند. روابط مربوط به آن در معادله ۵و ۶ آمده است. همین مفهوم را می‌توان به یک فضای جستجوی n بعدی تعمیم داد. در این حالت گرگ‌های خاکستری پیرامون بهترین راه حل به دست آمده در ابعادی بیشتر از ابعاد مکعب حرکت می‌کنند. همانطور که پیش تر گفته شد شکار بوسیله آلفا رهبری می‌شود.. به منظور 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 و تغییر دهند.

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

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

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

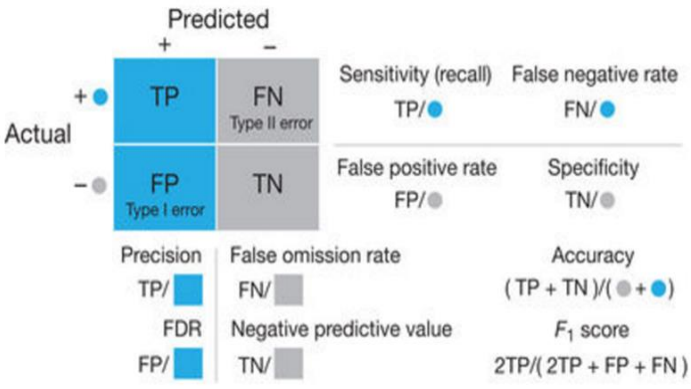
در این الگوریتم پیاده‌سازی فاز بهره‌برداری یا حمله که در صورت متوقف شدن طعمه اتفاق می‌افتد، با کاهش مقدار متغیر a از ۲ به ۱ انجام می‌شود. مقدار A نیز وابسته به a است، از این رو کاهش پیدا می‌کند. با کاهش مقدار A گرگ‌ها مجبور به حمله به سمت طعمه می‌شوند. همچنین برای جلوگیری از گیر افتادن در حداقل محلی، 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 تأثیرگذار دیگر بر فرآیند شناسایی مقدار C است. مقدار این بردار عددی تصادفی در بازه [۰،۲] است. این مقدار تصادفی تأثیر موقعیت طعمه را در تعیین فاصله، شدت (C>۱) یا ضعف (C<۱) می‌بخشد [۲۵].

۳،۳. فرآیند ارزیابی در سیستم پیشنهادی

در این بخش، معیارهای مختلفی که برای ارزیابی الگوریتم‌های دسته بندی وجود دارند، مورد بحث قرار خواهند گرفت. برای سادگی، تعریف این معیارها برای یک مسئله 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[۱] بیان می‌شود. پس از آن تعریف‌هایی از معیارهای مختلف ارزیابی مطرح خواهد شد.

مثبت کاذب[۲] 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.



[۱] confusion matrix

[۲] False Positive

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

منفی کاذب[۱] 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،

شکل ۲ ماتریس پراکندگی و معیار های ارزیابی (Lever et al, 2016)

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

مثبت واقعی[۲] 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.

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

منفی واقعی[۳] 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[۲۶].

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

۴،۳. ماتریس پراکندگی

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

دایره‌های آبی و خاکستری به ترتیب نشان دهنده‌ی موارد واقعا مثبت (TP + FN) و موارد واقعا منفی (FP + TN) است. مربع-های آبی و خاکستری نیز، به ترتیب موارد پیش‌بینی شده ی مثبت (TP + FP) و پیش‌بینی شده‌ی منفی (FN + TN) را نشان می-دهد. [۲۶].

۵،۳. معیارهای ارزیابی

از جمله مهم‌ترین معیارهایی که برای ارزیابی الگوریتم‌های دسته-بندی وجود دارند،

معیارهای

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Accuracy،

Precision،

Recall و F-measure 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، بدست می‌آید.

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

معیار Precision نشان‌دهنده نسبت کل موارد صحیح طبقه بندی شده در طبقه به کل موارد صحیح و غلط در طبقه می‌باشد.. این معیار به ما می‌گوید که چند درصد از جواب‌های درست، درست بوده است. به عبارتی صحت دسته‌بندی دسته i را با توجه به کل مواردی نشان. اندیس i در این پارامترها به این مفهوم است که پارامترها باید برای هر دسته i محاسبه شوند[۲۷].

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

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 می-باشد. نحوه محاسبه این معیار در ادامه آورده شده است.

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

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 پیش

[۱] False Negative
[۲] True Positive
[۳] True Negative

بینی کرده است و تا چه اندازه می‌توان به خروجی اعتماد کرد. [۲۹].

معیار F1، یک معیار مناسب برای ارزیابی دقت یک آزمایش (۱۵)

است. این معیار Precision و Recallرا با هم در نظر می‌گیرد. معیار F1 در بهترین حالت، یک و در بدترین حالت صفر است.

معیار Support یا پشتیبانی نشان دهنده 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 آنلاین بهره برده شده است. یکی از این بسترها، وب سایت شناخته شده اسنپ تریپ[۱] می‌باشد که در زیر مجموعه‌های خود، خدمات سفر و نقل و انتقالات را ارائه می‌دهد. به این ترتیب مجموعه نظرات سایت اسنپ تریپ به عنوان مجموعه داده مورد بررسی استفاده شده است همچنین در این پژوهش از زبان برنامه-نویسی پایتون و کتاب‌خانه‌های هوش مصنوعی مطرح در این زبان بهره‌برده شده است.

۴. ۲. آماده سازی داده

به دلیل آنکه متن‌ها اغلب شامل فرمت‌های خاصی که به متن کاوی کمک نمی‌کنند و می‌توانند حذف شوند، لازم است پیش از هرگونه تحلیل بر روی داده ها، بسته به نوع مسئله فرمت‌های غیرضروری حذف شوند. فرآیند پیش پردازش در این تحقیق از تکنیک‌هایی چون حذف کلمات ایست واژه ها، حذف تک کاراکتر ها، حذف فاصله‌های اضافی، حذف علامت‌های نگارشی، اعمال TFIDF و ایجاد ngram ها استفاده شده است. در فرآیند پیش پردازش متون انجام شده ابتدا ایست واژه ها[۲] حذف می‌شوند. ایست واژه ها 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 کارکتر ها و فاصله‌های اضافه‌ی موجود در متن‌ها و علامت‌های نگارشی حذف می‌شوند. پس از آن عملیات TFIDF برای تعیین میزان اهمیت کلمه ها 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 سایرِ متون کمتر تکرار شود(IDF) ، مقدار TF-IDFآن بیشتر می‌شود و این معیار خوبی جهت تشخیص وزنِ یک کلمه در یک جمله باشد. در واقع نشان می‌دهد که یک کلمه چقدر می‌تواند یکتا و مهم باشد.سپس n-gram ها تعیین می‌شود در مدلِ N-Gram برای ساختِ کوله‌ای از کلمات (BoW[۳])علاوه بر این‌که هر کلمه، یک ستونِ ویژگی (یا بُعد) هست، ترکیب کلمات نیز به ابعاد اضافه می‌شوند. در نهایت با توجه به نامتوازن بودن مجموعه داده ها در این تحقیق، داده‌ها متوازن خواهد شد.

۳،۴. اعمال روش ماشین بردار پشتیبان

در این بخش روش ماشین بردار پشتیبان بر روی مجموعه داده ی این تحقیق اعمال می‌شود. جدول۱ ، 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 داده ها، 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 می باشد. مقدار Accuracy این روش برابر با ۰,۶۹ درصد می‌باشد.

ماتریس درهم ریختگی به دست آمده حاصل از پیاده‌سازی ماشین بردار پشتیبان در شکل شماره۳ ارائه گردیده است. با توجه

[۱] snapptrip.com

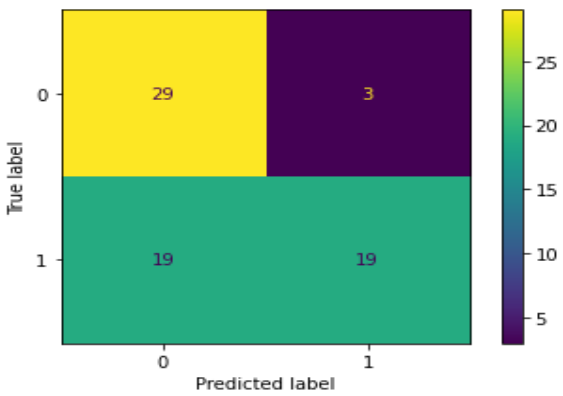
[۲] Stop Words

[۳] Bag of Word

به نتایج حاصله از این ماترس موارد به اشتباه تشخیص داده شده در دو کلاس به ترتیب برابر با ۱۹ و ۳ است.

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

	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



شکل ۳: ماتریس درهم ریختگی به دست آمده حاصل از پیاده‌سازی ماشین بردار پشتیبان

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 راه حل نشان دهنده‌ی تعدادی ویژگی است. برای ارزیابی راه حل مورد نظر، ماشین بردار پشتیبان برای ویژگی‌های انتخاب شده پیاده سازی می‌شود. در نهایت بهترین راه حل نشان دهنده‌ی ویژگی‌هایی است که مناسب ترین کارایی را دارند. در این بخش، با بهره‌گیری از روش گرگ-های خاکستری مناسب ترین ویژگی‌ها برای استفاده‌ی ماشین بردار پشتیبان تعیین می‌شود.

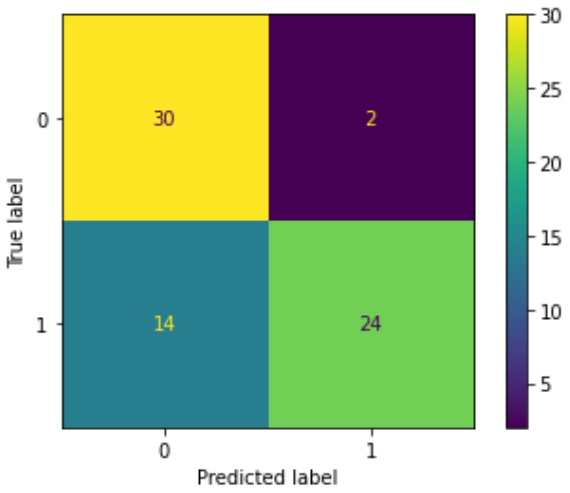
نتایج پیاده سازی روش ماشین بردار پشتیبان با روش گرگ خاکستری در جدول شماره ۲ ارائه گردیده است. از نتایج حاصل می توان دریافت که با اعمال سیستم پیشنهادی بر روی داده ها، مقادیر Precision برای کلاس اول و دوم برابر با ۰,۶۸ و ۰,۹۲ می‌باشد. مقادیر Recall نیز برای کلاس اول و دوم به ترتیب برابر با ۰,۹۴ و ۰,۶۳ می‌باشد. مقدار Accuracy این روش برابر با ۰,۷۷ درصد می‌باشد. مقایسه ی 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 پیاده سازی ماشین بردار پشتیبان و الگوریتم گرگ خاکستری

	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 درهم ریختگی به دست آمده حاصل از سیستم پیشنهادی برای عقیده کاوی در شکل شماره ۴ ارائه گردیده است. با توجه به خروجی حاصله از این ماترس موارد به اشتباه تشخیص داده شده در هر دو کلاس تعداد کم تری را نسبت به قبل نشان می‌دهند. این مسئله عملکرد مناسب سیستم پیشنهادی این تحقیق را در هر دو کلاس نشان می‌دهد



شکل ۴ : ماتریس درهم ریختگی به دست آمده حاصل از پیاده سازی 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*[۱] *[۱۷]*، روش ترکیبی شبکه¬ی¬عصبی حافظه¬ی طولانی کوتاه‌مدت *(LSTM)*[۲] و شبکه‌های عصبی کانولوشن *(CNN)*[۳] *[۸]*، روش ترکیبی بهینه¬سازی ازدحام ذرات *(PSO)*[۴] و جستجوی فاخته *(CS)*[۵] *[۱۰]*، و روش ترکیبی الگوریتم متاهیوریستیک *COBRA*[۶] و الگوریتم *SVM*[۷] *[۱۲]*، مقایسه گردید که نتایج در جدول شماره ۳ قابل رویت می‌باشد.

جدول ۳: مقایسه نتایج روش پیشنهادی.

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

همانگونه که از جدول مقایسه نتایج روش پیشنهادی با سایر روش

[۱] Lernt Independent Component Support Vector Regressive Deep

[۲] Long Short-term Memory Networks

[۳] Convolutional Neural Network

[۴] Particle Swarm Optimization

[۴] Cuckoo Search

[۴] Co-operation of Biology Related Algorithms

[۴] Support Vector Machine

ها پیداست، دقت بدست آمده از روش پیشنهادی، در قیاس با روش دوم(که در آن ویژگی‌های متن که از طریق *CNN*بدست آمده به عنوان ورودی برای شبکه‌های عصبی بازگشتی جهت تحلیل احساسات بکار برده شده) و روش سوم (که بر اساس یک الگوریتم *PSO* خود تطبیقی بهبود یافته با تابع تشخیص و جستجوی فاخته بوده است)و روش چهارم (که از روش های 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 روش های بررسی شده سبب افزایش دقت و کاهش زمان برای استخراج کلمات گردیده است می توان از الگوریتم گرگ خاکستری به دلیل کاهش ابعاد و افزایش سرعت همگرایی جهت تعیین ویژگی های مناسب پارامترها در *ICSD* استفاده نمود و بدین ترتیب بتوان به عملکرد بالاتری دست یافت.

۵. بحث و نتیجه‌گیری

امروزه اینترنت بشریت را تشویق می‌کند تا نظرات خود را از طریق وبلاگ‌ها، وب سایت ها و توییت ها 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 تکنیک های مختلفی توسط محققان در گذشته توسعه داده شده است. در این زمینه، رویکردهای مبتنی بر یادگیری ماشین به دلیل مکانیسم یادگیری طبقه‌بندی‌کننده‌ها، امیدوارکننده‌ترین تکنیک در نظر گرفته شده‌اند. در این کار، تکنیک ماشین بردار پشتیبان و رویکرد ترکیبی آن با الگوریتم بهینه‌سازی گرگ‌های خاکستری برای عقیده کاوی مورد بررسی قرار گرفت. نتایج پیاده سازی این

سیستم ها نشان داد که تلفیق روش های 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 به کار رفته است و عملکرد ماشین بردار پشتیبان را به نحو مطلوبی بهبود داد به گونه ای که عملکرد سیستم پیشنهاد در هر دو کلاس بهتر از گذشته شد.

منابع

[1] R.K. Bakshi. N. Kaur. R. Kaur. &, G. Kaur. **"**Opinion mining and sentiment analysis**"**. *3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. 2016**.** pp452-455.

[2] H. Chen. D. Zimbra. **"**AI and opinion mining**"**. *IEEE Intelligent Systems***.** *25*(3)**.** pp.74-80. 2010. https://doi.org/10.1109/MIS.2010.75

[۳] علی مردانی،س. آقایی،ع . "ارائه روش نظارتی برای نظرکاوی در زبان فارسی با استفاده از لغت نامه و الگوریتم SVM ".مدیریت فناوری اطلاعات.جلد ۷،شماره۲.ص۳۴۵-.۳۶۲,۱۳۹۴.

https://doi.org/10.22059/jitm.2015.53995

[4] M.Y. Raut. M.A. Kulkarni. "Polarity shift in opinion mining". In *2016 IEEE International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT)*. dec.2016. pp 333-337.

[5] R.S. Kumar. A.F. Saviour Devaraj. M Rajeswari. E.G. Julie. Y.H. "Robinson . Exploration of sentiment analysis and legitimate artistry for opinion mining". *Multimedia Tools and Applications*. pp.1-16. 2021. https://doi.org/10.1007/s11042-020-10480-w

[6] R. Ganesh. D. Saketh. VDV. S.K. M. Ramesh. "Website Evaluation Using Opinion Mining". In *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*. pp. 1499-1506. IEEE.sep.2021. DOI: 10.1109/ICIRCA51532.2021.9544566

[7] D. S. Abdelminaam, 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," in *IEEE Access*, vol. 9, pp. 97079-97099, 2021, doi: 10.1109/ACCESS.2021.3094173.

[8] H. Elzayady. K.M. Badran. G.I. Salama." Arabic Opinion Mining Using Combined CNN-LSTM Models". *International Journal of Intelligent Systems & Applications*. Vol.12(4), 2020. DOI: 10.5815/ijisa.2020.04.0

[9] M. Keyvanpour. Z. Karimi Zandian. M. Heidarypanah. "OMLML: a helpful opinion mining method based on lexicon and machine learning in social networks". *Social Network Analysis and Mining*, vol. 10(1). pp. 1-17. 2020. https://doi.org/10.1007/s13278-019-0622-6

[10] E. Souza. D Santos. G. Oliveira. A. Silva. A.L Oliveira. "Swarm optimization clustering methods for opinion mining". *Natural computing*, vol. *19*. pp.547-575.2020. https://doi.org/10.1007/s11047-018-9681-2

[11] E. Souza. D. Santos. G .Oliveira. A. Silva,and A.L. Oliveira . "Swarm optimization clustering methods for opinion mining". *Natural computing*. Vol. *19*. pp.547-575. 2020. https://doi.org/10.1007/s11047-018-9681-2

[12] A. Da'u. N. Salim. I. Rabiu . A. Osman. "Recommendation system exploiting aspect-based opinion mining with deep learning method". *Information Sciences*. Vol. *512*, pp.1279-12. 2020. https://doi.org/10.1016/j.ins.2019.10.038

[13] S. Akhmedova. E. Semenkin. V. Stanovov. "Co-operation of biology related algorithms for solving opinion mining problems by using different term weighting schemes". *Informatics in control, automation and robotics*. p. 73-90. 2018. https://doi.org/10.1007/978-3-319-55011-4_4

[14] B.V. Krishna. A.K. Pandey. A.S. Kumar. "Feature based opinion mining and sentiment analysis using fuzzy logic". *In: Gurumoorthy S, Rao BNK, Gao XZ (eds) Cognitive science and artifcial intelligence.* Singapore, pp 79–89. Springer. 2018. https://doi.org/10.1007/978-981-10-6698-6_8

[15] R. Narayan. J.K. Rout. S.K. Jena. "Review spam detection using opinion mining*". in Progress in intelligent computing techniques: Theory, practice, and applications.* pp. 273-279. Springer. 2018. DOI:https://doi.org/10.1007/978-981-10-3376-6_30

[16] K.S. Kumar. J. Desai. J. Majumdar. “Opinion mining and sentiment analysis on online customer review”. in 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). 2016. IEEE. DOI: 10.1109/ICCIC.2016.7919584

[17] S.R. El-Beltagy. T. Khalil. A. Halaby. M Hammad. “Combining lexical features and a supervised learning approach for Arabic sentiment analysis”. In *Computational Linguistics and Intelligent Text Processing: 17th International Conference, CICLing 2016, Konya, Turkey, April 3–9, 2016, Revised Selected Papers, Part II 17* .pp. 307-319. Springer International Publishing. 2018. https://doi.org/10.1007/978-3-319-75487-1_24

[18] MM. Krishna. B. Duraisamy. J. Vankara. “Independent component support vector regressive deep learning for sentiment classification”. *Measurement: Sensors*. 1;26. 100678.17. Apr 2023. https://doi.org/10.1016/j.measen.2023.100678

[**۱۹**]علیان نژادی.م.م.، حسینی.م. غضنفری.ح. کمندی.ا. "بازیابی تصویر مبتنی بر محتوا با استفاده از ماشین بردار پشتیبان و ویژگی‌های هیستوگرام اختلاف بافت"، *محاسبات نرم، جلد ۱۱، شماره ۱، ص. ۱۰-۲۱، ۱۴۰۱.*

https://doi.org/10.22052/scj.2022.246175.1053.

[**۲۰**] فیضی درخشی. م.، متقی نیا.ز.، عسگری چناقلو. م.، «طبقه‌بندی متون فارسی مبتنی بر شبکه‌های عصبی عمیق»، مجله محاسبات نرم، جلد ۱۱، شماره ۱، ص. ۸۴-۹۳، ۱۴۰۱. https://doi.org/10.22052/scj. 2023.243182.1010

[**۲۱**] زارع مهرجردی. ف.، یزدانیان دهکردی.م.، لطیف. ع.م.، «ارزیابی روش‌های یادگیری کلاسیک و یادگیری عمیق در تجزیه و تحلیل احساسات داده‌های تلگرام فارسی»، مجله محاسبات نرم، جلد ۱۱، شماره ۱، ۱۴۰۱. https://doi.org/10.22052/scj.2023.246553.10770

[22] T. Shaik. X. Tao. C. Dann. H. Xie., Y. Li L. Galligan. “Sentiment analysis and opinion mining on educational data: A survey”. *Natural Language Processing Journal*, pp.100003. 2022. https://doi.org/10.1016/j.nlp.2022.100003

[**۲۳**]راعی. ر. فلاح پور.س." کاربرد ماشین بردار پشتیبان در پیش‌بینی درماندگی مالی شرکت‌ها با استفاده از نسبت‌های مالی" . *بررسی‌های حسابداری و حسابرسی.جلد۱۵* ، شماره۱۳۸۷،۴.Dor:20.1001.1.26458020.1387.15.4.2.6

[24] K.-S. Shin. T.S. Lee. H.-j. Kim. “An application of support vector machines in bankruptcy prediction model”. *Expert systems with applications*. Vol. 28(1). pp. 127-135. 2005. https://doi.org/10.1016/j.eswa.2004.08.009

[25] S. Mirjalili. S.M. Mirjalili. A. Lewis. “Grey wolf optimizer”. *Advances in engineering software*. Vol. 69. pp . 46-61. 2014. https://doi.org/10.1016/j.advengsoft.2013.12.007

[26] A. Mohammadzadeh. M. Masdari. F. Soleimanian Gharehchopogh. A. Jafarian. “An improved grey wolves optimization algorithm for workflow scheduling in cloud computing environment”. *Journal of Soft Computing and Information Technology*, vol. *8*(4). pp.17-29. 2020.

[27] J. Lever. “Classification evaluation: It is important to understand both what a classification metric expresses and what it hides”. *Nature methods*, vol. 13(8). pp. 603-605. 2016.

[**۲۸**]**صنیعی آباده.م. محمودی. س. طاهرپور.م . "داده کاوی کاربردی". انتشارات نیاز دانش. ۱۳۹۱.**

[29] D.M. Powers. “Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation”. arXiv preprint arXiv:2010.16061. 2020. https://doi.org/10.48550/arXiv.2010.16061

[**۳۰**]خدایار. م. عصاره.ع. "نفوذ در شبکه های کامپیوتری با استفاده از تکنیک های ترکیب یادگیری ماشین". نهمین سمپوزیوم بین المللی پیشرفتهای علوم و تکنولوژی،مشهد،۱۳۹۳.

https://civilica.com/doc/841537.