هفتمین کنفرانس ملی مباحث نوین در کامپیوتر و فناوری اطلاعات
7th National Conference on Advanced Topics in Computer and Information Technology
۲ آذر ماه ۱۴۰۲

پیش‌بینی وقوع جرم و جنایت در جوامع با استفاده از داده کاوی

مهران اسماعیلی

چکیده : پیشگیری از جرم همواره یکی از موضوعات اساسی و مهم در زندگی بشری بوده که در طول تاریخ به شیوه‌های مختلف اعمال گردیده است. با توجه به گسترش فناوری‌های اطلاعات و ارتباط و راه‌اندازی سیستم‌های جامع اطلاعاتی در نیروی انتظامی و ثبت اطلاعات مجرمان در بانک‌های اطلاعاتی، کاربرد تکنیک‌های داده کاوی و کشف دانش به منظور تحلیل و ریشه‌یابی جرایم از جمله سرقت یکی از ضرورت‌های پلیس و دستگاه‌های قضایی می‌باشد. لذا در این تحقیق سعی در استفاده از الگوریتم‌های داده کاوی و زنبورعسل به منظور پیش‌بینی و تشخیص وقوع جرم و جنایت در جوامع را داریم. نتایج بدست آمده از پیاده‌سازی روش پیشنهادی حاکی از آن است که توانسته است با دقت ۹۷.۷۷ درصد برای داده‌های آموزشی و دقت ۹۳.۸۱ درصد برای داده‌های آزمایشی به پیش‌بینی و تشخیص وقوع جرم و جنایت در جوامع بپردازد.

کلمات کلیدی: پیش‌بینی– جرم و جنایت - داده کاوی- الگوریتم زنبور عسل

۱ - مقدمه

پیشگیری از جرم همواره یکی از موضوعات اساسی و مهم در زندگی بشری بوده که در طول تاریخ به شیوه‌های مختلف اعمال گردیده است. با توجه به گسترش فناوری‌های اطلاعات و ارتباط و راه‌اندازی سیستم‌های جامع اطلاعاتی در نیروی انتظامی و ثبت اطلاعات مجرمان در بانک‌های اطلاعاتی، کاربرد تکنیک‌های داده کاوی و کشف دانش به منظور تحلیل و ریشه‌یابی جرایم از جمله سرقت یکی از ضرورت‌های پلیس و دستگاه‌های قضایی می‌باشد[۶,۵]. جوامع چندین سال است که با معضل جرم و چگونگی کنترل آن درگیر بوده‌اند که تمرکز و خط مشی اصلی آنان در کنترل جرم تنها افزایش و حرفه‌ای‌تر کردن سیستم پلیسی و قضایی بوده است. ایجاد امنیت و آرامش در جامعه تنها با توسل به شیوه‌های کیفری پس از وقوع جرم محقق نمی‌شود بلکه دولت وظیفه دارد با در پیش گرفتن راه‌کارهایی قبل از وقوع جرم در از بین بردن شرایط تحقق آن تلاش نماید[۸,۷]. بررسی تحلیل و ریشه‌یابی جرائم مستلزم بهره‌گیری از ابزارها و تکنیک‌های ارائه شده در حوزه فناوری اطلاعات می‌باشد. پیشرفت‌های اخیر فناوری در پردازش و تجزیه و تحلیل داده‌های بزرگ و پیچیده باعث ارائه فرصت‌ها و چالش‌هایی برای پلیس و سازمان‌های امنیتی شده است که با استفاده از داده‌های عظیم و متنوع می‌توان در سه حوزه کلیدی پیشگیری از جرم، تشخیص جرم و جنایت و امنیت ملی پرداخت؛ همچنین مسائل نظارتی و درک عمومی در مورد حفظ حریم خصوصی آزادی‌های مدنی و مزایای اجتماعی آن را نیز پوشش می‌دهد[۱۰,۹]. در بین فناوری‌های معرفی شده داده کاوی به عنوان یک ابزار قدرتمند در تحلیل داده‌های جرم نقش مؤثری در کشف الگو و دانش از پایگاه داده جرم در راستای فرایند تصمیم‌گیری به منظور کاهش فرصت ارتکاب جرم و کنترل جرایم دارد[۱۱]. در برخی کشورها از الگوریتم‌ها و تکنیک‌های داده کاوی و ابزارهای آن به منظور تحلیل جرائم استفاده می‌شوند همچنین به دلیل عدم وجود سیستم مکانیزه و ثبت اطلاعات مجرمین در سال‌های گذشته برای کشف جرائم تنها از کارآگاهان و افسران ماهر و با تجربه استفاده می‌شد[۱۳,۱۲].

۲– عوامل مؤثر در وقوع جرم

عوامل مؤثر در ایجاد رفتار مجرمانه شامل موارد بسیاری می‌شود. در این میان می‌توان به محیط دوران کودکی، توارث، موقعیت اجتماعی، موقعیت اقتصادی، رویدادهای بحرانی و منفی اشاره کرد[۱۸].

- عوامل اقتصادی موثر در وقوع جرم و جنایت
- فقر
- نابرابری درآمدی

- نرخ بیکاری
- صنعتی شدن و شهرنشینی
- تجربه گذشته فرد
- تورم
- سطح آموزش

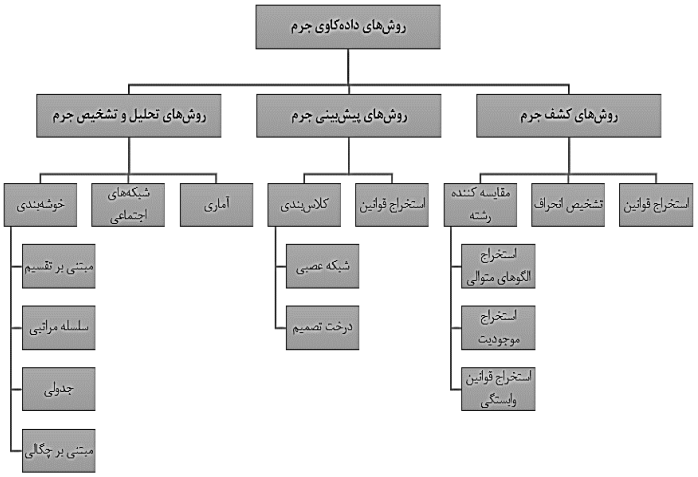
۳– داده کاوی جرم

امروزه به دلیل وجود ابزارهای مختلف برای جمع‌آوری داده‌ها و پیشرفت قابل قبول تکنولوژی پایگاه داده حجم انبوهی از اطلاعات در انبار داده‌های مختلف ذخیره شده است، این رشد انفجاری داده‌ها احتیاج به یک سری تکنیک‌ها و ابزارهای جدید که توانایی پردازش هوشمندانه اطلاعات را دارا باشند نمایان می‌سازد. داده کاوی با پیدا کردن مجموعه‌ای از الگوهای جالب از دل داده‌های موجود در انباره‌ها قادر است چنین نیازی را مرتفع کند. در حال حاضر داده کاوی توسط بسیاری از محققان به عنوان یک موضوع تحقیقاتی مهم به شمار می‌رود. محققان در بسیاری از رشته‌ها نظیر پایگاه داده‌ها، یادگیری ماشین و آمار این موضوع را پیگیری کرده و تکنیک‌های مختلفی را برای داده کاوی، تکنیک‌ها و روش‌های مختلف ارائه شده در این زمینه را معرفی کرده و آنها را طبقه‌بندی کند. داده کاوی یکی از مهمترین مراحل فرآیند استخراج دانش در پایگاه داده به حساب می‌آید. الگوریتم‌های داده کاوی جرم در حقیقت سازوکاری برای ایجاد یک مدل از درون بانک‌های اطلاعاتی مربوط به جرائم می‌باشد. برای ایجاد یک مدل در ابتدا یک الگوریتم باید مجموعه‌ای از داده‌ها را تحلیل کرده و یک الگوی استخراج را تعریف کند. کشف دانش در پایگاه داده فرایند شناسایی درست، ساده، مفید و در نهایت الگوها و مدل‌های قابل فهم در داده‌ها است که داده کاوی مرحله‌ای از فرآیند کشف دانش است و شامل الگوریتم‌های مخصوص به خود است بطوری که تحت محدودیت‌های مؤثر محاسباتی قابل قبول، الگوها و مدل‌ها را در داده‌ها کشف می‌کند؛ منظور از الگوی مفید مدلی در داده‌ها است که ارتباط میان زیر مجموعه‌ای از داده‌ها را توصیف می‌کند در حالی که معتبر، ساده، قابل فهم و جدید است. در واقع داده کاوی یکی از مهمترین روش‌هایی است که بوسیله آن الگوهای مفید در درون داده‌ها با حداقل دخالت کاربران شناخته و اطلاعاتی را در اختیار کاربران و تحلیلگران قرار می‌دهند تا براساس آنها تصمیمات مهم و حیاتی در سازمان‌ها اتخاذ شود[۲۴,۲۳].

انتخاب درست یک الگوریتم برای استفاده در یک حادثه و یا جرم اتفاق افتاده می‌تواند یک چالش باشد و ممکن است استفاده از الگوریتم‌های مختلف برای اجرای همان کار لازم باشد. بطورکلی روش‌ها و الگوریتم‌های داده کاوی به سه دسته کلی تقسیم می‌شوند[۲۶,۲۵] :

۱. اکتشاف جرم
۲. پیشگویی و پیش‌بینی جرم
۳. آنالیز و تحلیل جرم

انتخاب درست یک روش و یا الگوریتم برای یک هدف خاص از داده کاوی جرم توسط کاربر و باتوجه به نیاز و هدف کاری او می‌تواند تعیین شود. در شکل (۱) روش‌های مورد استفاده در داده کاوی جرائم نشان داده شده است.



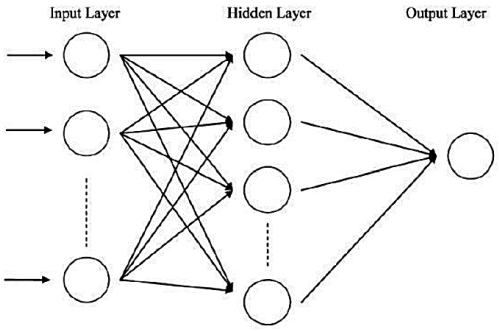
شکل (۱) روش‌های مورد استفاده در داده کاوی جرائم

۴– شبکه عصبی مصنوعی

شبکه‌های عصبی مدل ساده شده‌ای است که بر مبنای عملکرد مغز انسان کار می‌کند. اساس کار این شبکه شبیه‌سازی تعداد زیادی واحد پردازشی کوچک است که با هم در ارتباط هستند. به هریک از این واحدها یک نرون[1] گفته می‌شود. نرون‌ها بصورت لایه لایه قرار دارند و در یک شبکه عصبی معمولا سه لایه وجود دارد. اولین لایه، لایه ورودی نام دارد، دومین لایه، لایه نهان بوده و سومین لایه، لایه خروجی است که لایه نهان می‌تواند متشکل از یک لایه یا بیشتر باشد [۲۷]. شبکه عصبی مصنوعی

[1] Neuron

که توسط شبیه‌سازی مغز انسان توسعه یافته است یک ابزار مدلسازی خطی انعطاف‌پذیر است و توانایی یادگیری انسان در یک محیط کامپیوتری با شبکه عصبی مصنوعی منتقل می‌شود. شبکه عصبی مصنوعی از تعدادی از عناصر پردازشی در چارچوب قوانین خاصی تشکیل شده است که نرون‌ها یا گره‌ها نامیده می‌شوند. یک شبکه عصبی مصنوعی به طور کلی شامل سه لایه از نرون‌های متصل به هم است. اولین لایه، لایه ورودی نامیده می‌شود. هر نرون در لایه ورودی سیگنال‌ها را به لایه پنهان می‌فرستد. اطلاعات دریافت شده از لایه ورودی در لایه پنهان پردازش می‌شود. اطلاعات خروجی به لایه خروجی منتقل می‌شود. در شکل (۲) سه لایه مربوط به شبکه عصبی مصنوعی نشان داده شده است:



شکل (۲) شبکه عصبی مصنوعی[۲۷]

۵- انتخاب ویژگی

انتخاب ویژگی یکی از روش‌های کاهش بعد است که به استخراج تنها یک زیرمجموعه از ویژگی‌ها می‌پردازد و نوع خاص از ویژگی‌ها وجود دارند[۲۹,۲۸]:

- **ویژگی‌های اضافی:** ویژگی‌هایی که شامل اطلاعات موجود در یک یا چند ویژگی دیگر هستند. مانند ویژگی سن در حالی که ویژگی تاریخ تولد هم وجود دارد.
- **ویژگی‌های نامرتبط:** ویژگی‌هایی که حاوی هیچ اطلاعات مفیدی برای آن وظیفه داده کاوی نیستند. مانند شماره ردیف در رده‌بندی درختان که این ویژگی‌ها می‌توانند دقت پیش‌بینی و ثبات را کاهش دهند با انتخاب زیرمجموعه مناسب این ویژگی‌ها حذف می‌شوند و در فرآیند داده کاوی به کار گرفته نمی‌شوند.

روش‌های انتخاب ویژگی به سه دسته تقسیم می‌شوند که عبارتند از :

۵-۱- روش‌های فیلتر[۲]

این روش‌ها از تکنیک رتبه‌بندی ویژگی‌ها به همراه یک معیار برای دسته‌بندی ویژگی‌ها استفاده می‌کند، برای هر ویژگی یک رتبه یا امتیاز محاسبه شده و سپس براساس این رتبه ویژگی‌ها دسته‌بندی می‌شوند و ویژگی‌های با امتیاز کمتر حذف می‌شوند. در این روش از یک آستانه برای رتبه‌بندی ویژگی‌ها استفاده می‌شود. این روش‌ها بر ویژگی‌های کلی مجموعه داده آموزش تکیه دارند و فرآیند انتخاب ویژگی را به عنوان یک گام پیش‌پردازش با استقلال از الگوریتم استقرایی انجام می‌دهند. مزیت این مدل‌ها هزینه محاسباتی پایین و توانایی تعمیم خوب آنها محسوب می‌شود.

۵-۲- روش‌های پوشاننده[۳]

در روش‌های پوشاننده عملکرد یک الگوریتم یادگیری برای ارزیابی زیرمجموعه ویژگی‌ها مورد استفاده قرار می‌گیرد به عبارت دیگر الگوریتم انتخاب ویژگی از روش یادگیری به عنوان یک زیرمجموعه با بار محاسباتی استفاده می‌کند که از فراخوانی الگوریتم برای ارزیابی هر زیرمجموعه از ویژگی‌ها نشئت می‌گیرد. با این حال این تعامل با دسته‌بند منجر به نتایج کارایی بهتری نسبت به فیلترها می‌شود. در این روش‌ها به منظور تعیین میزان خوب بودن هر یک از زیرمجموعه ویژگی‌های تولید شده از الگوریتم‌های یادگیری استفاده می‌شود. در روش‌های پوشاننده از یک تابع طبقه‌بندی برای ارزیابی شایستگی زیرمجموعه‌های ویژگی استفاده می‌شود. زیرمجموعه‌ها توسط یک الگوریتم جستجو، کشف و اعتبارسنجی می‌شوند. این روش‌ها به دلیل استفاده از طبقه‌بند دارای پیچیدگی محاسباتی بسیار زیاد و قابلیت تعمیم کمتر در مقایسه با روش‌های فیلتری هستند که این امر سبب محدودیت استفاده از آنها در داده‌های با ابعاد بالا می‌شود.

۵-۳- روش‌های جداسازی

در روش‌های جاسازی شده انتخاب ویژگی را در فرآیند آموزش انجام می‌دهند و معمولا برای ماشین‌های یادگیری خاصی مورد استفاده قرار می‌گیرند. در این روش‌ها جستجو برای یک زیرمجموعه بهینه از ویژگی‌ها

[2] Filter methods

[3] Wrappers methods

در مرحله ساخت طبقه‌بند انجام می‌شود و می‌توان آن را بعنوان جستجویی در فضای ترکیبی از زیرمجموعه‌ها و فرضیه‌ها دید. این روش‌ها قادر به ثبت وابستگی‌ها با هزینه‌های محاسباتی پایین‌تر نسبت به طبقه‌بندها هستند. هریک از روش‌های فیلتر و پوشاننده مزایا و معایبی دارند و به عبارتی مکمل یکدیگر خواهند بود. بنابراین روش‌های جاسازی شده با بکارگیری مزیت‌های هر دو روش جواب‌های بهتری تولید می‌کنند.

۶– الگوریتم زنبور عسل

الگوریتم کلونی مصنوعی زنبورعسل در سال ۲۰۰۹ توسط Karaboga ارائه گردید که از شبیه‌سازی رفتار جستجوی غذای گروه‌های زنبورعسل ایجاد شده است[۳۱]. در این الگوریتم، متغیرهای مسئله به عنوان موقعیت منبع غذایی در نظر گرفته می‌شوند که مقدار شهد[۴] منابع غذایی نشان دهنده تابع احتمال و یا برازندگی هر راهحل می‌باشد. هر منبع غذایی توسط یک زنبور کارگر[۵] کشف و استخراج می‌شود به عبارت دیگر می‌توان گفت تعداد زنبورهای کارگر برابر تعداد منابع غذایی پیرامون کندو[۶] خواهد بود.

۷ – کارهای پیشین

در ادامه و در جدول (۱) خلاصه‌ای از کارهای انجام شده در زمینه تشخیص جرم و جنایت در جوامع آمده است.

جدول (۱) خلاصه کارهای پیشین

بهترین نتیجه بدست آمده	مرجع	روش استفاده شده
الگوریتم رگرسیون با دقت ۱۰۰ درصد	[۳۲]	الگوریتم‌های رگرسیون و درخت تصمیم‌گیری
الگوریتم جنگل تصادفی با دقت ۸۳.۳۹ درصد	[۳۳]	الگوریتم‌های درخت تصمیم‌گیری، جنگل تصادفی، ناوبیز و رگرسیون خطی
الگوریتم درخت تصمیم با دقت ۱۰۰ درصد	[۳۴]	الگوریتم‌های درخت تصمیم، نظریه بیزین، شبکه پرسپترون چندلایه و بردار ماشین پشتیبان
استخراج ویژگی بر مبنای	[۳۵]	استفاده از چهار روش انتخاب
همبستگی FS4 با دقت ۹۵.۹۴ درصد		ویژگی به صورت FS1 تا FS4
استفاده از مدل KNN با دقت ۹۳,۹۹ درصد	[۳۶]	استفاده از تکنیک‌های داده کاوی

۸– شرح روش پیشنهادی

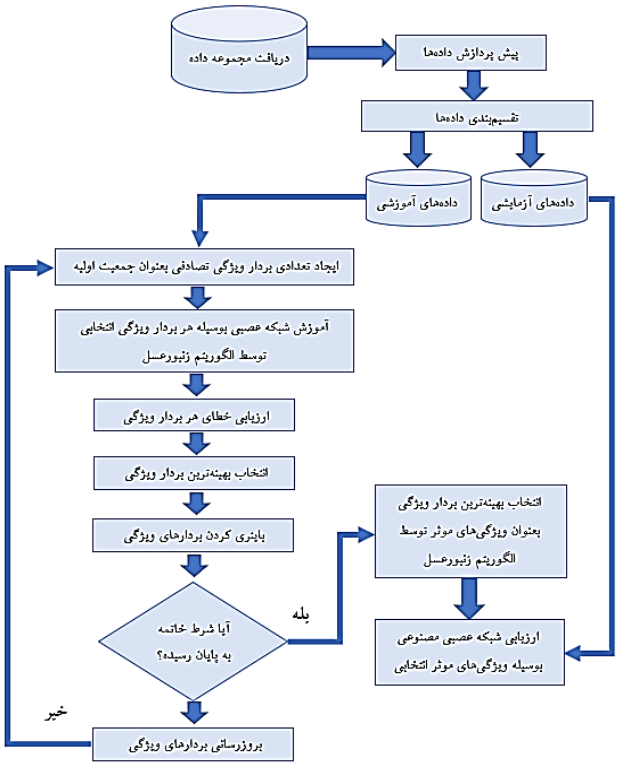
در این تحقیق با استفاده از الگوریتم زنبورعسل جهت انتخاب ویژگی‌های برتر و استفاده از شبکه عصبی مصنوعی به پیش‌بینی و تشخیص وقوع جرم و جنایت در جوامع خواهیم پرداخت. در روش پیشنهادی برای آنکه دقت شبکه عصبی مصنوعی افزایش داده شود یک بردار از ویژگی‌های مرتبط با وقوع جرم و جنایت در جوامع به عنوان یک عضو الگوریتم زنبورعسل در نظر گرفته می‌شود. در روش پیشنهادی برای آنکه یادگیری شبکه عصبی با خطای کمتری مواجه شود همزمان با آموزش شبکه عصبی مصنوعی تلاش می‌شود تا ویژگی‌های مهم برای پیش‌بینی وقوع جرم و جنایت در جوامع توسط الگوریتم زنبورعسل انتخاب شود، بنابراین مراحل روش پیشنهادی به صورت زیر است :

- داده‌های مرتبط با وقوع جرم و جنایت در جوامع به عنوان ورودی در نظر گرفته می‌شوند
- داده‌ها مورد پیش پردازش قرار گرفته شده و در ادامه این داده‌ها به دو بخش داده‌های آموزشی و آزمایشی تقسیم می‌شوند
- یک بردار ویژگی به عنوان یک عضو الگوریتم زنبورعسل کدگذاری می‌شود به نحوی که این بردار ویژگی دارای مقادیر صفر و یک است و بیانگر عدم انتخاب ویژگی و انتخاب ویژگی می‌باشند
- تعدادی از بردارهای ویژگی به عنوان جمعیت اولیه الگوریتم زنبورعسل تصادفی ایجاد می‌شوند و هرکدام از مولفه‌های آن می‌تواند دارای مقدار صفر و یک باشد؛ هر بردار ویژگی که دارای مقدار صفر باشد یعنی این ویژگی برای یادگیری در شبکه عصبی مصنوعی مهم نبوده و ویژگی‌هایی که برابر یک هستند برای یادگیری در شبکه عصبی مصنوعی مهم هستند
- هر بردار ویژگی یک عضو الگوریتم زنبورعسل بوده که توسط تابع هدف ارزیابی می‌شود

[4] Nectar
[5] Employed Bee
[6] Hive

- تابع هدف بیانگر میزان خطای شبکه عصبی مصنوعی می‌باشد
- هر بردار ویژگی یا یک عضو الگوریتم زنبورعسل که بتواند تابع هدف را کمیته کند به عنوان بردار ویژگی بهینه و موثر انتخاب می‌شود.
- در هر تکرار الگوریتم زنبورعسل یک واحد به شمارنده اضافه می‌شود و در ادامه می‌توان تعداد تکرار را در هر بار اجرا چک کرد و اگر از تعداد بیشینه تکرار کمتر بود مراحل الگوریتم پیشنهادی تکرار می‌شود و اگر از مقدار بیشینه تکرار بیشتر بود آنگاه مراحل تکرار تمام شده و بردار ویژگی بهینه برای یادگیری شبکه عصبی مصنوعی استفاده می‌شود.

در ادامه و در شکل (۳) با توجه به توضیحات بیان شده فلوچارت روش پیشنهادی نشان داده شده است.



شکل (۳) فلوچارت روش پیشنهادی

همانطور که مشاهده می‌شود پس از دریافت مجموعه داده، پیش‌پردازش بر روی داده‌ها صورت گرفته و در ادامه به تقسیم‌بندی داده‌ها پرداخته و شبکه عصبی مصنوعی ایجاد می‌شود. پس از ایجاد شبکه عصبی مصنوعی ویژگی‌های مفید از بین تمامی ویژگی‌های موجود در مجموعه داده با استفاده از الگوریتم زنبورعسل انتخاب شده و شبکه عصبی مصنوعی با استفاده از زیرمجموعه ویژگی‌های انتخاب شده مورد آموزش قرار گرفته می‌شود. در پایان بعد از بررسی شرط خاتمه در صورتی که به شرط خاتمه رسیده باشیم روش پیشنهادی مورد ارزیابی قرار گرفته می‌شود اما در صورتی که به شرط خاتمه نرسیده باشیم شبکه عصبی مجددا مورد آموزش قرار گرفته می‌شود.

۹- ارزیابی روش پیشنهادی

در این بخش به بررسی انواع معیارهای مورد استفاده در این تحقیق جهت سنجش و ارزیابی روش پیشنهادی جهت پیش‌بینی و تشخیص وقوع جرم و جنایت در جوامع با استفاده از ماتریس آشفتگی خواهیم پرداخت. ماتریس آشفتگی چگونگی عملکرد الگوریتم طبقه‌بندی را با توجه به مجموعه داده ورودی نشان می‌دهد که شامل چهار حالت بوده و هرکدام از این حالات مفهومی ویژه دارند، در شکل (۴) ماتریس آشفتگی نشان داده شده است.

		برچسب واقعی	
		کلاس مثبت (وقوع جرم و جنایت)	کلاس منفی (عدم وقوع جرم و جنایت)
برچسب پیش‌بینی توسط روش پیشنهادی	کلاس مثبت (وقوع جرم و جنایت)	TP	FP
	کلاس منفی (عدم وقوع جرم و جنایت)	FN	TN

شکل (۴) ماتریس آشفتگی

از جمله معیارهایی که با توجه به ماتریس آشفتگی جهت سنجش و ارزیابی کارایی روش پیشنهادی در این تحقیق می‌توان استفاده نمود معیارهای دقت، فراخوانی، صحت و F می‌باشند که در معیارهای مذکور مفاهیم TP تعداد درستی‌های مثبت، TN تعداد درستی‌های منفی، FP تعداد خطاهای مثبت و FN تعداد خطاهای منفی در تشخیص وقوع جرم و جنایت صورت گرفته در جوامع می‌باشند.

$$Accuracy = \left(\frac{TP+TN}{TP+TN+FP+FN}\right) \times 100 \quad (۱)$$

$$Recall = \left(\frac{TP}{TP+FN}\right) \times 100 \quad (۲)$$

$$Precision = \left(\frac{TP}{TP+FP}\right) \times 100 \quad (۳)$$

$$F = \frac{2 \times Recall \times precision}{Recall+precision} \quad (۴)$$

۱۰- پیاده سازی روش پیشنهادی

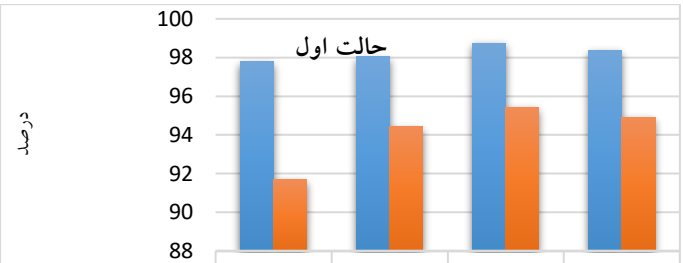
در این تحقیق سعی کردیم با استفاده از الگوریتم زنبورعسل به انتخاب زیرمجموعه ویژگی‌های موثر جهت پیش‌بینی و تشخیص وقوع جرم و جنایت در جوامع بپردازیم. برای این منظور پس از دریافت مجموعه داده مورد نظر از سایت UCI به اعمال پیش‌پردازش بر روی داده‌های آن پرداختیم. بنابراین ۲۷ ویژگی را که دارای مقادیر زیادی داده گمشده بودند از مجموعه داده حذف نمودیم در نتیجه تعداد ویژگی‌ها به ۱۰۰ ویژگی کاهش یافت و در ادامه داده‌ها را در بازه بین صفر و یک مورد نرمالسازی قرار دادیم. پس از اعمال پیش‌پردازش به تقسیم‌بندی داده‌ها به دو دسته آموزشی و آزمایشی پرداخته شد و در ادامه پس از ساخت شبکه عصبی مصنوعی با استفاده از الگوریتم زنبورعسل به انتخاب ویژگی‌های موثر پرداخته و در ادامه روش پیشنهادی بوسیله ویژگی‌های موثر انتخابی مورد آموزش و ارزیابی قرار گرفت. در جدول (۲) مشخصات و پارامترهای مورد استفاده در این تحقیق آمده است :

جدول (۲) پارامترهای مورد استفاده

پارامترها	مقادیر پارامترها
تعداد لایه پنهان	یک لایه
تعداد نورون لایه پنهان	۳۵ نورون
تعداد منابع	براساس تعداد ویژگی‌های موثر انتخابی
جمعیت زنبورها	۵۰ زنبور
تعداد تکرار	۱۰۰۰

۱۱- ارزیابی نتایج

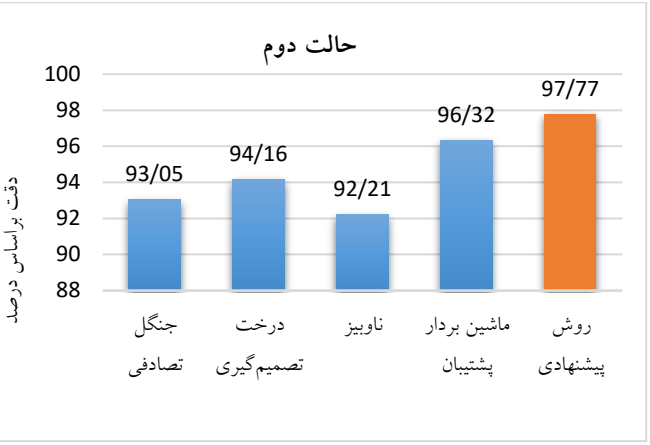
در این بخش به ارزیابی روش پیشنهادی در دو حالت خواهیم پرداخت. در حالت اول نتایج بدست آمده از روش پیشنهادی را با نتایج بدست آمده در حالتی که از تمامی ویژگی‌های مجموعه داده استفاده شده است مورد مقایسه قرار دادیم و در حالت دوم بجای استفاده از طبقه‌بند شبکه عصبی مصنوعی از دیگر طبقه‌بندهای داده‌کاوی جهت ارزیابی ویژگی‌های انتخابی استفاده نمودیم.



	دقت	صحت	فراخوانی	F
روش پیشنهادی	97/77	98/03	98/75	98/38
بدون انتخاب ویژگی	91/69	94/43	95/43	94/92

شکل (۵) مقایسه نتایج بدست آمده در حالت اول

باتوجه به نتایج بدست آمده و همانگونه که در شکل (۵) نیز نشان داده شده است دقت بدست آمده برای داده‌های آموزشی در روش پیشنهادی همراه با انتخاب ویژگی نسبت به حالتی که از انتخاب ویژگی استفاده نشده است ۶.۰۸ درصد افزایش داشته است همچنین مقادیر بدست آمده از معیارهای صحت، فراخوانی و F در روش پیشنهادی همراه با انتخاب ویژگی به ترتیب ۳.۶ درصد، ۲.۶ درصد و ۳.۴۶ درصد نسبت به حالتی که از انتخاب ویژگی استفاده نشده است می‌باشد. در نتیجه می‌توان گفت روش پیشنهادی همراه با انتخاب ویژگی عملکرد بهتری را جهت پیش‌بینی و تشخیص وقوع جرم و جنایت در جوامع نشان می‌دهد. در ادامه و در حالت دوم روش پیشنهادی خود را با برخی دیگر از طبقه‌بندهای داده کاوی مورد مقایسه قرار دادیم بدین منظور بجای شبکه عصبی مصنوعی از طبقه‌بندهایی همچون ماشین بردار پشتیبان، ناوبیز، درخت تصمیم‌گیری و جنگل تصادفی استفاده نمودیم که نتایج آن در شکل (۶) مشاهده می‌شود.



شکل (۶) مقایسه دقت روش پیشنهادی با سایر طبقه‌بندهای داده کاوی در حالت دوم

باتوجه به نتایج بدست آمده و همانگونه که در شکل ۴-۴ نیز نشان داده شده است دقت بدست آمده در روش پیشنهادی نسبت به ماشین بردار پشتیبان ۱.۴۵ درصد، نسبت به ناوبیز ۵.۵۶ درصد، نسبت به درخت تصمیم‌گیری ۳.۶۱ درصد و نسبت به جنگل تصادفی ۴.۷۲ درصد افزایش داشته است. در نتیجه می‌توان گفت روش پیشنهادی نسبت به الگوریتم‌های مورد مقایسه از دقت و عملکرد بهتری جهت پیش‌بینی و تشخیص وقوع جرم و جنایت در جوامع برخوردار است.

۱۲- نتیجه گیری

پیشرفت‌های اخیر فناوری در پردازش و تجزیه و تحلیل داده‌های بزرگ و پیچیده باعث ارائه فرصت‌ها و چالش‌هایی برای پلیس و سازمان‌های امنیتی شده است که با استفاده از داده‌های عظیم و متنوع می‌توان در سه حوزه کلیدی پیشگیری از جرم، تشخیص جرم و جنایت و امنیت ملی پرداخت؛ همچنین مسائل نظارتی و درک عمومی در مورد حفظ حریم خصوصی آزادی‌های مدنی و مزایای اجتماعی آن را نیز پوشش می‌دهد. در بین فناوری‌های معرفی شده داده کاوی به عنوان یک ابزار قدرتمند در تحلیل داده‌های جرم نقش مؤثری در کشف الگو و دانش از پایگاه داده جرم در راستای فرایند تصمیم‌گیری به منظور کاهش فرصت ارتکاب جرم و کنترل جرایم دارد. لذا در این تحقیق سعی کردیم با استفاده از الگوریتم زنبورعسل به انتخاب زیرمجموعه ویژگی‌های موثر جهت پیش‌بینی و تشخیص وقوع جرم و جنایت در جوامع با دقت بالا بپردازیم. برای این منظور پس از دریافت مجموعه داده مورد نظر و اعمال پیش‌پردازش بر روی داده‌های آن به تقسیم‌بندی داده‌ها به دو دسته آموزشی و آزمایشی پرداخته شد و در ادامه پس از ساخت شبکه عصبی مصنوعی، با استفاده از الگوریتم زنبورعسل ویژگی‌های موثر را انتخاب کرده و به آموزش و ارزیابی روش پیشنهادی بوسیله ویژگی‌های انتخابی پرداخته شد. نتایج بدست آمده از پیاده‌سازی روش پیشنهادی حاکی از آن است که توانسته است با دقت ۹۷.۷۷ درصد برای داده‌های آموزشی و دقت ۹۳.۸۱ درصد برای داده‌های آزمایشی به پیش‌بینی و تشخیص وقوع جرم و جنایت در جوامع بپردازد؛ بنابراین دقت بدست آمده برای داده‌های آموزشی در روش پیشنهادی ۶.۰۸ درصد و برای معیارهای صحت، فراخوانی و F به ترتیب ۳.۶ درصد، ۲.۶ درصد و ۳.۴۶ درصد بهبود داشته است. در ادامه روش پیشنهادی خود را با برخی دیگر از طبقه‌بندهای داده کاوی مورد مقایسه قرار دادیم بدین منظور بجای شبکه عصبی مصنوعی از طبقه‌بندهایی همچون ماشین بردار پشتیبان، ناوبیز، درخت تصمیم‌گیری و جنگل تصادفی استفاده نمودیم که نتایج بدست آمده حاکی از آن است که دقت بدست آمده در روش پیشنهادی نسبت به ماشین بردار پشتیبان ۱.۴۵ درصد، نسبت به ناوبیز ۵.۵۶ درصد، نسبت به درخت تصمیم‌گیری ۳.۶۱ درصد و نسبت به جنگل تصادفی ۴.۷۲ درصد افزایش داشته است. در نتیجه می‌توان گفت روش پیشنهادی نسبت به الگوریتم‌های مورد مقایسه از دقت و عملکرد بهتری جهت پیش‌بینی و تشخیص وقوع جرم و جنایت در جوامع برخوردار است.

مراجع

[1] Oatley, G. C. (2022). Themes in data mining, big data, and crime analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *12*(2), e1432.

[2] Das, D., & Nayak, M. (2021). Crime Pattern Detection Using Data Mining. *Intelligent Data Analytics for Terror Threat Prediction: Architectures, Methodologies, Techniques and Applications*, 221-236.

[3] Ferooz, F., Hassan, M. T., Awan, M. J., Nobanee, H., Kamal, M., Yasin, A., & Zain, A. M. (2021). Suicide bomb attack identification and analytics through data mining techniques. *Electronics*, *10*(19), 2398.

[4] Raza, D. M., & Victor, D. B. (2021, March). Data mining and Region Prediction Based on Crime

Economic Planning Sciences, *80*, 101043.

[15] Adhikary, A., Murad, S. A., Munir, M. S., & Hong, C. S. (2022, January). Edge assisted crime prediction and evaluation framework for machine learning algorithms. In *2022 International Conference on Information Networking (ICOIN)* (pp. 417-422). IEEE.

[16] Kumar, S., Ahmed, R., Bharany, S., Shuaib, M., Ahmad, T., Tag Eldin, E., ... & Shafiq, M. (2022). Exploitation of Machine Learning Algorithms for Detecting Financial Crimes Based on Customers' Behavior. *Sustainability*, *14*(21), 13875.

[17] Singh, A., Srivastava, A. P., Bhardwaj, G., Jain, A., & Singh, A. K. (2022, April). Methods to Detect an Event using Artificial Intelligence and Machine Learning. In *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)* (pp. 297-301). IEEE.

[18] Karabogas. D And Basturk.B, (2009), A Powerful And Efficient Algorithm For Numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm, Journal Of Global Optimization 39, pp. 459-471.

[19] McClendon, L., & Meghanathan, N. (2021). Using machine learning algorithms to analyze crime data. *Machine Learning and Applications: An International Journal (MLAIJ)*, *2*(1), 1-12.

[20] Yerpude, P. (2020). Predictive modelling of crime data set using data mining. *International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol*, *7*.

[21] Ivan, N., Ahishakiye, E., Omulo, E. O., & Wario, R. (2019). A performance analysis of business intelligence techniques on crime prediction.

[22] Mohd, J., & Noor, N. M. M. (2017). A comparative study to evaluate filtering methods for crime data feature selection. *Procedia computer science*, *116*, 113-120.

[23] Tayal, D. K., Jain, A., Arora, S., Agarwal, S., Gupta, T., & Tyagi, N. (2015). Crime detection and criminal identification in India using data mining techniques. *AI & society*, *30*(1), 117-127.

Using Random Forest. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)* (pp. 980-987). IEEE.

[5] Hamdi, A., Shaban, K., Erradi, A., Mohamed, A., Rumi, S. K., & Salim, F. D. (2022). Spatiotemporal data mining: a survey on challenges and open problems. *Artificial Intelligence Review*, *55*(2), 1441-1488.

[6] Al-Ghushami, A. H., Syed, D., Sessa, J., & Zainab, A. (2022, June). Intelligent Automation of Crime Prediction using Data Mining. In *2022 IEEE 31st International Symposium on Industrial Electronics (ISIE)* (pp. 245-252). IEEE.

[7] Raja, R. A., Yuvaraj, N., & Kousik, N. V. (2021). Analyses on Artificial Intelligence Framework to Detect Crime Pattern. *Intelligent Data Analytics for Terror Threat Prediction: Architectures, Methodologies, Techniques and Applications*, 119-132.

[8] Babu, S. Z. D., Pandey, D., & Sheik, I. (2020). An overview of a crime detection system using the art of data mining. *International Journal of Innovations in Engineering Research and Technology*, *7*(05), 125-128.

[9] Jabeen, N., & Agarwal, P. (2021). Data mining in crime analysis. In *Proceedings of Second International Conference on Smart Energy and Communication* (pp. 97-103). Springer, Singapore.

[10] Patil, R., Kacchi, M., Gavali, P., & Pimparia, K. (2020). Crime pattern detection, analysis and prediction using machine learning. *International Research Journal of Engineering and Technology (IRJET)*, *ISSN*, 2395-0056.

[11] Gupta, M. K., & Chandra, P. (2020). A comprehensive survey of data mining. *International Journal of Information Technology*, *12*(4), 1243-1257.

[12] Saravanan, P., Selvaprabu, J., Arun Raj, L., Abdul Azeez Khan, A., & Javubar Sathick, K. (2021). Survey on crime analysis and prediction using data mining and machine learning techniques. In *Advances in Smart Grid Technology: Select Proceedings of PECCON 2019—Volume II* (pp. 435-448). Springer Singapore.

[13] Bandekar, S. R., & Vijayalakshmi, C. (2020). Design and analysis of machine learning algorithms for the reduction of crime rates in India. *Procedia Computer Science*, *172*, 122-127.

[14] Wang, H., & Ma, S. (2022). Preventing crimes against public health with artificial intelligence and machine learning capabilities. *Socio-*