یازدهمین همایش سالانه
بانکداری نوین و
نظام‌های پرداخت
پژوهشکده پولی و بانکی
بانک مرکزی جمهوری اسلامی ایران
بانک مرکزی جمهوری اسلامی ایران
شرکت ملی انفورماتیک
زیست بوم بانکداری هوشمند؛ حکمرانی دیجیتال و نظارت فنّاورانه
تهران، مرکز همایش‌های بین‌المللی برج میلاد ـ ۵ و ۶ آذر ۱۴۰۳ ISC

بهبود یادگیری ماشین در تشخیص تقلب کارت‌های اعتباری

امین عارف[۱] **فرشید فرح خواه**[۲]

چکیده

تشخیص تقلب در سیستم‌های بانکی برای ثبات مالی، حفاظت از مشتری و انطباق با مقررات بسیار مهم است. یادگیری ماشینی در بهبود تجزیه‌وتحلیل داده‌ها و همچنین تشخیص تقلب در زمان واقعی حیاتی است. انتخاب ویژگی به‌عنوان یک فاز مهم در یادگیری ماشین برای افزایش تشخیص تقلب کارت اعتباری[۳] ضروری است. درواقع از بین بردن تأثیر منفی ویژگی‌های اضافی و نامربوط و انتخاب ویژگی‌های مؤثر در فاز انتخاب ویژگی، به فاز طبقه‌بندی در یادگیری ماشین کمک می‌کند. در این مقاله یک روش مؤثر بر اساس الگوریتم بهینه‌سازی یوزپلنگ[۴] برای افزایش ظرفیت شناسایی دقیق تراکنش‌های کلاهبرداری کارت اعتباری با شناخت ویژگی‌های مربوطه ارائه شده است. همچنین در این تحقیق در فاز طبقه‌بندی یادگیری ماشین، روش ماشین بردار پشتیبان قرار دارد که از طریق تنظیم پارامترها با استفاده از الگوریتم یوزپلنگ بهبود یافته است. نتایج نشان داده است که روش پیشنهادی بروی مجموعه داده اعتبار استرالیا نسبت به روش‌های دیگر حداقل ۱/۴۸ درصد بهبود در تشخیص تقلب به همراه داشته است.

واژه‌های کلیدی: تشخیص تقلب کارت اعتباری، یادگیری ماشین، الگوریتم بهینه‌سازی یوزپلنگ
طبقه‌بندی JEL: G21، E58، E51

[۱] کارشناس گروه امنیت و ریسک شرکت شاپرک A.aref@shaparak.com
[۲] رییس گروه امنیت و ریسک شرکت شاپرک F.farahkhah@shaparak.com

[3] Credit card fraud detection

[4] The cheetah optimizer

۱ مقدمه

تجارت الکترونیک[1] به معاملات دیجیتالی کالا یا خدمات اشاره دارد و امکان خرید محصولات را برای مصرف‌کنندگان تسهیل کرده است [۱، ۲]. چهارچوب‌های تجارت الکترونیک مزایای متعددی از جمله تسریع روند تدارکات، کاهش هزینه‌ها، افزایش راحتی مصرف‌کننده، تسهیل مقایسه محصول و قیمت، انطباق سریع با نوسانات بازار و روش‌های پرداخت مختلف را ارائه می‌کنند. این جنبه‌ها به انعطاف‌پذیری اقتصادی کمک کرده‌اند، به‌ویژه در طول محدودیت‌های دولتی ناشی از بیماری همه‌گیر و دستورات حبس خانگی، مانند مواردی که در دوره‌های همه‌گیری کرونا تجربه شده‌اند [۳،۴].

قبل از بحران مالی ۲۰۰۸ ـ ۲۰۰۹، رشد سالانه درآمدهای تجارت الکترونیک از ۱۵٪ به ۲۵٪ افزایش یافت اما به دلیل رکود اقتصادی به حدود ۳٪ در سال ۲۰۰۹ کاهش یافت [۵]. این نرخ رشد به طور قابل‌توجهی پایدارتر از کاهش تجربه شده در تجارت خرده‌فروشی سنتی در همان دوره بود. پس از سال ۲۰۰۹، نرخ‌های رشد تجارت الکترونیک افزایش یافت و سالانه بیش از ۱۰ درصد افزایش یافت که نرخی بسیار بالاتر از کل فروش خرده‌فروشی بود. همه‌گیری کرونا در سال ۲۰۲۰، در حالی که بر فروش دیجیتالی تأثیر منفی می‌گذارد، همچنین منجر به افزایش قابل‌توجه در درآمد تجارت الکترونیک خرده‌فروشی شد، که به دلیل تغییر مصرف‌کننده به خرید آنلاین، روندی که پس از همه‌گیری ادامه داشت، تاکنون دنبال شده است [۵]. همه‌گیری مداوم کرونا تقاضای آنلاین برای محصولات مختلف از جمله اقلام ضروری را افزایش داد. این امر همچنین استفاده از روش‌های پرداخت الکترونیکی را افزایش داده و باعث افزایش قابل توجه فعالیت‌های مالی متقلبانه شده است [۶].

با توجه به تکامل تجارت الکترونیک، تقریباً همه مشاغل، چه در مقیاس کوچک و چه در مقیاس بزرگ، پرداخت با کارت اعتباری را پذیرفته‌اند. با این حال، این افزایش استفاده از کارت اعتباری، به ویژه از طریق تراکنش‌های آنلاین، فرصت‌های جدیدی را برای مجرمان برای سوءاستفاده از اطلاعات کارت اعتباری مصرف‌کننده فراهم کرده است. کلاهبرداری مالی با پیامدهای گسترده خود، چالشی جدی در حوزه‌های آکادمیک، تجاری و نظارتی ایجاد می‌کند و بر ارائه‌دهندگان خدمات و مشتریان آنها تأثیر مخربی می‌گذارد. این موضوع در عرصه مالی بسیار مهم است و در تعاملات اقتصادی روزمره در مقیاس جهانی نفوذ می‌کند. این امر مستلزم تصاحب غیرمجاز دارایی‌ها یا سرمایه برای خود غنی‌سازی است که اعتماد به نهادهای مالی را کاهش می‌دهد و هزینه زندگی را افزایش می‌دهد. طبقه‌بندی کلاهبرداری مالی چندوجهی است و شامل جعل صورت‌های مالی، فریب در شیوه‌های بیمه، اختلاس بانکی، کلاهبرداری از راه دور، و فعالیت‌های متقلبانه در بازارهای سهام و کالا می‌شود [۷،۸]. اثرات موج‌دار چنین رفتار متقلبانه‌ای، تحولی قابل توجه در زیرساخت‌های مالی جهانی ایجاد کرده است، و انتقال به سمت خدمات مالی دیجیتالی را تسریع کرده و عصر جدیدی از چالش‌ها را آغاز نموده است.

---

[1] E-commerce

گزارش‌ها حاکی از افزایش شدید زیان‌های کلاهبرداری مرتبط با کارت‌های اعتباری و بدهی است که از سال ۲۰۰۰ تا ۲۰۱۵ به طور تصاعدی رشد کرده است [۹] و نشان می‌دهد که خریدهای غیرمجاز و کارت‌های اعتباری جعلی، اگرچه تنها ۱۰ تا ۱۵ درصد هستند ولی از مجموع موارد کلاهبرداری، مسئول ۷۵ تا ۸۰ درصد خسارات مالی ناشی از کلاهبرداری است. در نتیجه، بخش‌های خصوصی و دولتی سرمایه‌گذاری‌های خود را در تحقیق و توسعه افزایش می‌دهند تا سیستم‌های پیچیده‌تری برای کشف تقلب ایجاد کنند. جریان مالی قابل توجه در تجارت الکترونیکی فعالیت‌های متقلبانه را جذب می‌کند و خطر زیان‌های مالی قابل‌توجهی را به همراه دارد. تا اینجا که افزایش تقلب از ۱۷/۵ میلیارد دلار در سال ۲۰۲۳ به ۲۰ میلیارد دلار در سال ۲۰۲۴ گزارش شده است. این ارقام بر ضرورت توسعه مکانیسم‌های شناسایی و پیشگیری از تقلب برای تجارت الکترونیک و نهادهای مالی تأکید می‌کند [۱۰]. کلاهبرداری یک عمل نامشروع است که با فریب مشخص می‌شود. کلاهبرداری کارت اعتباری[1] مستلزم به دست آوردن غیرقانونی اطلاعات دارنده کارت از طریق ارتباط تلفنی، پیام‌های متنی یا هک سایبری برای تسهیل تراکنش‌های مالی غیرمجاز است. چنین اعمال متقلبانه اغلب با استفاده از نرم‌افزار تحت کنترل مجرم انجام می‌شود. روش شناسایی کلاهبرداری کارت اعتباری زمانی شروع می‌شود که یک مصرف‌کننده تراکنش را با استفاده از اعتبار خود انجام دهد و این تراکنش باید برای تایید مشروعیت آن تایید شود [۱۱ـ۱۲].

تلاش‌ها در تحقیقات به سمت توسعه سیستم‌های تشخیصی است که از روش‌هایی مانند یادگیری ماشینی، یادگیری عمیق و داده‌کاوی استفاده می‌کنند. این سیستم‌ها داده‌های تراکنش را برای تمایز بین فعالیت‌های معتبر و متقلبانه بررسی می‌کنند. پیچیدگی تشخیص کلاهبرداری کارت اعتباری در حال افزایش است زیرا ویژگی‌های تراکنش‌های متقلبانه به معاملات قانونی شبیه‌تر می‌شوند. در نتیجه، شرکت‌های کارت اعتباری مجبور به پیاده‌سازی فن‌آوری‌های پیشرفته‌تر کشف تقلب هستند. یک سیستم تشخیص تقلب کارآمد برای شناسایی دقیق فعالیت‌های تقلبی در تراکنش‌های بلادرنگ ضروری است که به طور گسترده به سیستم‌های تشخیص ناهنجاری و شناسایی سوءاستفاده طبقه‌بندی می‌شوند [۱۱]. مؤسسات مالی که کارت‌های اعتباری صادر می‌کنند یا تراکنش‌های آنلاین را مدیریت می‌کنند باید از سیستم‌های خودکار کشف تقلب استفاده کنند. این باعث کاهش ضرر و افزایش اعتماد مشتری می‌شود. با ظهور کلان داده‌ها و هوش مصنوعی، امکانات جدیدی برای استفاده از مدل‌های پیشرفته یادگیری ماشین برای کشف تقلب پدید آمده است [۱۳].

سیستم‌های تشخیص تقلب امروزی که بر روش‌های پیشرفته یادگیری ماشین متکی هستند، بسیار مؤثر هستند. یک مدل طبقه‌بندی باینری که می‌تواند بین تراکنش‌های عادی و متقلبانه تمایز قائل شود، با استفاده از مجموعه داده‌ای ساخته می‌شود که شامل تراکنش‌های برچسب‌گذاری شده (عادی و متقلبانه) است. مدل توسعه‌یافته تعیین می‌کند که آیا تراکنش‌های ورودی مشروع یا تقلبی هستند. شناسایی تراکنش‌های تقلبی با استفاده از تکنیک‌های طبقه‌بندی چالش‌های زیادی را به همراه دارد [۱۴،۱۵]. سیستم‌های خودکار برای کشف تقلب برای نهادهای مالی درگیر در صدور کارت‌های اعتباری یا رسیدگی به تراکنش‌های آنلاین ضروری هستند، زیرا به حداقل رساندن ضرر و تقویت اعتماد مشتری کمک می‌کنند. ظهور داده‌های بزرگ و هوش مصنوعی، فرصت‌های جدیدی را در استفاده

[1] Credit card fraud (CCF)

از مدل‌های پیچیده یادگیری ماشین برای کشف تقلب باز کرده است. به‌عنوان نمونه در تحقیق [۱۳]، نشان داده است که جدیدترین سیستم‌های کشف تقلب که از تکنیک‌های پیشرفته یادگیری ماشین استفاده می‌کنند و در این حوزه بسیار مؤثر هستند.

تحقیق حاضر به بررسی کلاهبرداری کارت اعتباری، زیرمجموعه مهمی از کلاهبرداری بانکی می‌پردازد. کارت‌های اعتباری، یک روش پرداخت الکترونیکی غالب در سراسر جهان، تراکنش‌های آنلاین را ساده کرده و فعالیت‌های کلاهبرداری مجرمان سایبری را افزایش داده‌اند. استفاده غیرمجاز از سیستم‌ها یا اطلاعات کارت اعتباری، اغلب بدون اطلاع مالک، یک نگرانی فزاینده است که بانک‌ها و نهادهای مالی بسیاری را در سطح جهان تحت تأثیر قرار می‌دهد. تشخیص کلاهبرداری کارت اعتباری شامل تجزیه‌وتحلیل خودکار داده‌های تراکنش ذخیره شده در پایگاه داده‌های ارائه‌دهنده خدمات است. انواع متداول کلاهبرداری شامل کلاهبرداری برنامه و تراکنش است که حوزه‌های فیزیکی و دیجیتالی را در بر می‌گیرد. مقوله‌های کلاهبرداری شامل موقعیت‌های بدون کارت، کارت‌های تقلبی و سرقت هویت می‌شود.

بانک‌ها نقش اساسی در بررسی تراکنش‌ها برای تقلب قبل از مجوز پرداخت دارند، به ویژه با بررسی اینکه آیا وب‌سایت درگیر در لیست بلوک قرار دارد یا خیر. تراکنش‌های مرتبط با سایت‌های مسدود شده به عنوان یک اقدام پیشگیری از کلاهبرداری رد می‌شوند. پس از تأیید بانک، فرآیند پیشگیری از تقلب صرفاً بر کشف تقلب تمرکز می‌کند. اقدامات پیشگیرانه برای تقلب در برنامه شامل اعتبار سنجی جزئیات شخصی است. ابزارهایی مانند خدمات تأیید آدرس، ارزش تأیید کارت، امنیت سه بعدی و رمزگذاری برای تقلب در تراکنش استفاده می‌شود. اهمیت تعداد ویژگی‌ها در مجموعه داده‌ها یک موضوع مهم است زیرا آنها می‌توانند اثربخشی روش‌های طبقه‌بندی را مختل کنند و به طور بالقوه باعث ایجاد بیش از حد یا ابهام در الگوریتم‌ها شوند [۱۶]. انتخاب ویژگی مستلزم شناسایی ویژگی‌های مؤثر و حذف ویژگی‌های اضافی است که بر عملکرد مدل یادگیری ماشین تأثیر می‌گذارد. هدف انتخاب ویژگی کاهش ابعاد مجموعه ویژگی‌ها در عین حفظ دقت عملکرد است.

روش‌های مختلفی برای طبقه‌بندی مجموعه داده‌ها توسعه داده شده‌اند، اما روش‌های فراابتکاری توجه قابل‌توجهی را به دلیل اثربخشی آنها در رسیدگی به طیف گسترده‌ای از مسائل بهینه‌سازی به خود جلب کرده است [۱۷]. روش‌های فراابتکاری تکنیک‌های بهینه‌سازی هستند که برای یافتن راه‌حل‌های تقریبی برای مسائل مختلف بهینه‌سازی طراحی شده‌اند. این الگوریتم‌ها جهت بهینه‌سازی نیازی به محاسبه مشتق ندارند و توانایی دور زدن بهینه محلی و رسیدن به بهینه سراسری مسائل را دارند. روش‌های فراابتکاری به صورت تصادفی عمل می‌کنند و برخلاف روش‌های جستجوی گرادیان که نیاز به محاسبه مشتقات در فضای جستجو دارند، فرآیند بهینه‌سازی خود را با راه‌حل‌های تصادفی تولید شده آغاز می‌کنند. سادگی و صریح بودن آنها، ناشی از مفاهیم اولیه و اجرای آسان، این الگوریتم‌ها را بسیار انعطاف‌پذیر و به راحتی قابل تغییر برای پاسخگویی به مسائل خاص می‌کند. ویژگی کلیدی روش‌های فراابتکاری ظرفیت استثنایی آنها برای جلوگیری از همگرایی زودرس است. ماهیت تصادفی آنها به آنها اجازه می‌دهد تا به طور مؤثر به عنوان یک «جعبه سیاه» عمل کنند، بهینه محلی را دور بزنند و فضای جستجو را به طور کامل کاوش کنند [۱۸ ـ ۲۳].

استفاده از روش‌های فراابتکاری برای رسیدگی به مسائل انتخاب ویژگی می‌تواند به طور مؤثر چالش‌های مختلف در تجزیه‌وتحلیل داده‌ها را برطرف کند. انتخاب ویژگی برای استخراج ویژگی‌های مرتبط از مجموعه داده‌های نامتعادل، به ویژه قبل از طبقه‌بندی موارد کلاهبرداری کارت اعتباری در مجموعه داده‌های گسترده، بسیار مهم است. مزایای اولیه انتخاب ویژگی شامل تفسیر ساده داده‌ها، کاهش مدت زمان آموزش و حل مسائل مربوط به ابعاد بالا است. الگوریتم‌های الهام‌گرفته از زیست‌شناسی، که در مقابله با مشکلات پیچیده و ترکیبی ماهر هستند، به طور مؤثر در تشخیص کلاهبرداری کارت اعتباری استفاده شده‌اند. قابل توجه در این میان، الگوریتم بهینه‌سازی یوزپلنگ است، یک روش بهینه‌سازی مشهور با الهام از زیستی که به دلیل قابلیت دوگانه جستجوی محلی و سراسری از طریق تکنیک‌های اکتشاف و بهره‌برداری شناخته شده است [۲۴ـ۲۷]. این مقاله از الگوریتم بهینه‌سازی یوزپلنگ به دلیل جستجوی جامع آن در توابع بهینه‌سازی در ابعاد بالا برای انتخاب ویژگی استفاده می‌شود و همچنین برای بهبود ماشین بردار پشتیبان استفاده شده است.

۲ مرور ادبیات

تحقیقات گسترده‌ای برای کشف تقلب مالی انجام شده است که تعداد زیادی از روش‌های پیشنهادی را شامل می‌شود که به طور گسترده در ادبیات فهرست‌بندی شده است (به عنوان مثال، در جدول ۱ اشاره شده است). این روش‌ها عمدتاً تکنیک‌های یادگیری ماشین هستند. هدف این مطالعه بررسی محدودیت‌های ذاتی روش‌های کلاس موجود، با تمرکز بر تکنیک‌های یادگیری ماشین است.

جدول ۱

مروری بر روش‌های یادگیری ماشین در تشخیص تقلب

مرجع	سال	مجموعه داده	روش	دستاورد
[۲۸]	۲۰۱۸	کارت‌خوان اروپا[1]	نزدیک‌ترین همسایه، بیزین	دقت ۹۷/۹۲
[۲۹]	۲۰۱۹	kaggle	شبکه عصبی، بیزین و ماشین بردار پشتیبان	دقت ۹۹/۹۳
[۳۰]	۲۰۱۹	kaggle	جنگل تصادفی	دقت ۹۹
[۳۱]	۲۰۱۹	برنامه داده‌های متریک تسهیلات[2]	انتخاب ویژگی و رگرسیون لجستیک	دقت ۸۶/۹۸
[۳۲]	۲۰۱۹	داده‌های کارت اعتباری[3]	شبکه عصبی و ماشین بردار پشتیبان	دقت ۹۶
[۳۳]	۲۰۲۱	داده برزیل و اروپا	شبکه عصبی LSTM	دقت ۷۷
[۳۴]	۲۰۲۱	کارت‌خوان اروپا	درخت تصمیم	دقت ۹۷/۱۸
[۳۵]	۲۰۲۱	سه مجموعه داده UCI	انتخاب ویژگی الگوریتم ژنتیک و شبکه عصبی، ماشین بردار پشتیبان	دقت ۸۱/۹۷
[۳۶]	۲۰۲۱	داده‌های استرالیا	شبکه عصبی خودسازمانده	دقت ۹۰
[۳۷]	۲۰۲۲	کارت‌خوان اروپا	ماشین بردار پشتیبان	دقت ۹۹

[1] European cardholders

[2] Facility Metrics Data Program

[3] Credit card datasets

[۷]	۲۰۲۲	اعتباری استرالیا	انتخاب ویژگی الگوریتم کرم شب‌تاب و ماشین بردار پشتیبان	دقت ۸۵/۶۵
[۳۸]	۲۰۲۲	کارت‌خوان اروپا	انتخاب ویژگی و ماشین بردار پشتیبان	دقت ۹۸/۶۰
[۳۹]	۲۰۲۲	کارت‌خوان اروپا	درخت تصمیم، نزدیک‌ترین همسایه و بیزین	دقت ۹۱/۱۱
[۱۵]	۲۰۲۲	کارت‌خوان اروپا	درخت تصمیم، جنگل تصادفی	دقت ۹۹/۸۰
[۴۰]	۲۰۲۲	kaggle	افزایش نمونه‌ها و روش جنگل تصادفی	دقت ۹۴
[۴۱]	۲۰۲۳	کارت‌خوان اروپا	افزایش نمونه‌ها و روش XGBoost	دقت ۹۹/۹۵
[۴۲]	۲۰۲۳	داده‌های بیمه خسارت پزشکی	خوشه‌بندی و روش رگرسیون لجستیک	دقت ۸۸
[۴۳]	۲۰۲۳	داده بنچمارک AAER	انتخاب ویژگی و بهبود طبقه‌بندی با الگوریتم مرغ ماهیگیر	دقت ۹۴
[۴۴]	۲۰۲۳	باشگاه وام‌دهی	جنگل تصادفی و رگرسیون لجستیک و ماشین بردار پشتیبان	دقت ۹۵
[۴۵]	۲۰۲۴	kaggle	جنگل تصادفی، رگرسیون لجستیک	دقت ۹۴
[۴۶]	۲۰۲۴	Sparkov	یادگیری عمیق اتوانکودر	دقت ۹۷
[۴۷]	۲۰۲۴	داده دنیای واقعی	یادگیری چندتایی	دقت ۹۹
[۴۸]	۲۰۲۴	اعتباری استرالیا	انتخاب ویژگی با الگوریتم خرس قهوه‌ای و ماشین بردار پشتیبان	دقت ۹۲

با توجه به مقالات بررسی شده مشخص است که مجموعه داده‌های مختلفی برای کشف تقلب ارائه شده است و روش‌های یادگیری ماشین با فاز انتخاب ویژگی و طبقه‌بندی در این حوزه توسعه یافته است. در آخرین تحقیقات این حوزه از جمله [۴۸] از انتخاب ویژگی با الگوریتم فراابتکاری خرس قهوه‌ای[1] و ماشین بردار پشتیبان[2] استفاده شده است. در مقاله حاضر از الگوریتم یوزپلنگ در انتخاب ویژگی[3] و همچنین بهبود ماشین بردار پشتیبان در مقایسه با مرجع [۴۸] استفاده شده است.

۳ روش پیشنهادی

یکی از چالش‌های اصلی در تشخیص تقلب، حجم عظیم اطلاعات و ویژگی‌های موجود در هر تراکنش مالی است. برای مثال، یک تراکنش می‌تواند شامل اطلاعات مختلفی مانند مبلغ، مکان جغرافیایی، زمان تراکنش، و نوع کالا باشد. با این حال، همه این اطلاعات برای شناسایی تقلب مهم نیستند. برخی ویژگی‌ها می‌توانند گیج‌کننده باشند و تأثیر منفی بر دقت مدل بگذارند.

اینجا جایی است که الگوریتم بهینه‌سازی یوزپلنگ وارد می‌شود. این الگوریتم با الهام از رفتار شکار یوزپلنگ، ویژگی‌های مهم و مرتبط را پیدا می‌کند. یوزپلنگ هنگام شکار، ابتدا به دقت محیط خود را بررسی می‌کند، سپس تصمیم می‌گیرد که بهترین طعمه (ویژگی) را انتخاب کند و با حداکثر سرعت به سمت آن حرکت می‌کند. در اینجا،

[1] brown bear optimization algorithm

[2] Support vector machine

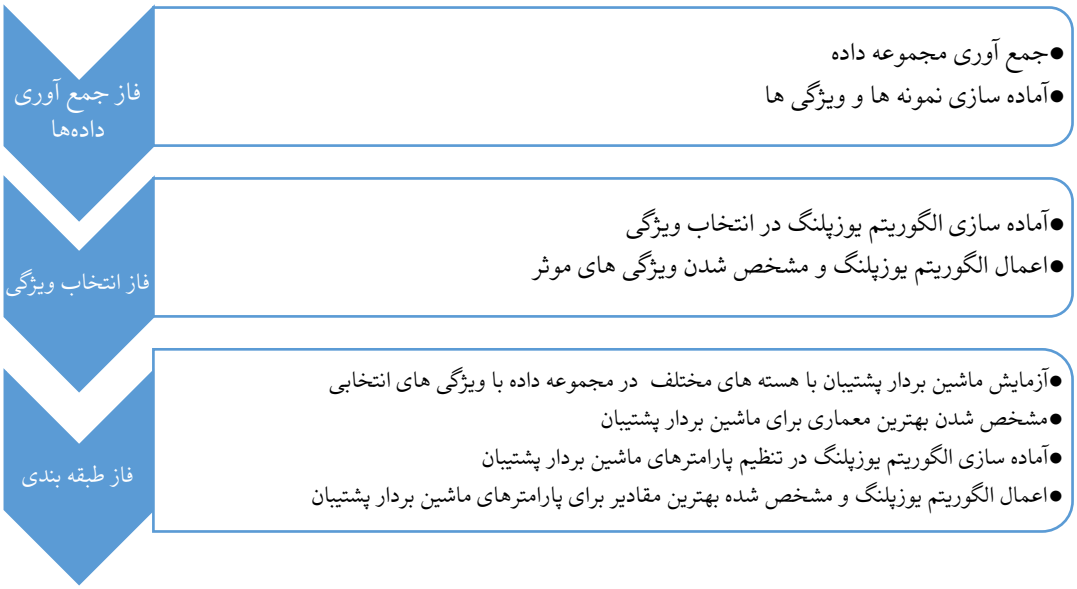
[3] Feature Selection

الگوریتم یوزپلنگ به سیستم کمک می‌کند تا ویژگی‌هایی که برای تشخیص تقلب مهم‌تر هستند را پیدا کند و ویژگی‌های غیرضروری یا اضافی را نادیده بگیرد.

برای مثال، اگر داده‌های تراکنش شامل ۱۰ ویژگی باشد، الگوریتم یوزپلنگ می‌تواند مشخص کند که از این ۱۰ ویژگی، مثلاً ۴ ویژگی بیشترین تأثیر را در شناسایی تقلب دارند. سپس این ۴ ویژگی به عنوان ورودی به مرحله بعد یعنی طبقه‌بندی فرستاده می‌شوند.

از چالش‌های روش پیشنهادی این است که چگونه می‌توان تشخیص تقلب را با دقت بالاتری انجام داد و چگونه روش ماشین بردار پشتیبان بهبود یافته به کمک الگوریتم بهینه‌ساز یوزپلنگ، در تشخیص تقلب می‌تواند نسبت به روش ماشین بردار پشتیبان استاندارد [۴۸] و نسخه بهبودیافته آن با الگوریتم خرس قهوه‌ای دقت بالاتری داشته باشد.

چارچوب روش تحقیق در روش پیشنهادی شامل سه قسمت است که در شکل ۱ آمده است.



شکل ۱ . فازهای روش تحقیق

همانطور که از شکل ۱ مشخص است فازهای روش تحقیق در مدل پیشنهادی شامل سه فاز است:

۱) فاز جمع‌آوری داده‌ها: در این فاز، مجموعه داده‌ها اعتبار استرالیا [49] آماده‌سازی می‌شود.

در این مرحله، داده‌های مورد نیاز برای آموزش مدل از یک مجموعه داده واقعی با نام مجموعه داده اعتبار استرالیا استخراج می‌شود. این داده‌ها شامل اطلاعات مختلفی از تراکنش‌های کارت اعتباری است؛ مانند مقدار پول در هر تراکنش، نوع کالا یا خدمات خریداری‌شده، زمان انجام تراکنش و سایر ویژگی‌ها.

هدف از این مرحله این است که یک بانک اطلاعاتی معتبر از تراکنش‌های قانونی و تقلبی فراهم شود. این داده‌ها به دو قسمت تقسیم می‌شوند:

- داده‌های آموزشی: برای آموزش مدل به کار می‌رود.
- داده‌های آزمایشی: برای بررسی عملکرد مدل پس از آموزش استفاده می‌شود.

در واقع، مدل ابتدا از داده‌های آموزشی یاد می‌گیرد که چگونه بین تراکنش‌های قانونی و تقلبی تفاوت قائل شود. سپس، با استفاده از داده‌های آزمایشی، عملکرد مدل ارزیابی می‌شود تا مشخص شود که چقدر خوب عمل می‌کند.

۲) فاز انتخاب ویژگی: در این فاز، ویژگی‌های مؤثر در مجموعه داده انتخاب می‌شود.

در هر تراکنش، اطلاعات زیادی وجود دارد که همه آن‌ها برای تشخیص تقلب مفید نیستند. بعضی ویژگی‌ها اهمیت بیشتری دارند و به شناسایی تقلب کمک می‌کنند، در حالی که برخی دیگر ممکن است غیرضروری یا حتی باعث کاهش دقت تشخیص شوند. بنابراین، نیاز است که فقط ویژگی‌های مرتبط و مهم برای تشخیص تقلب انتخاب شوند.

اینجاست که الگوریتم بهینه‌سازی یوزپلنگ وارد عمل می‌شود. این الگوریتم با الهام از رفتار شکار یوزپلنگ طراحی شده است. همان‌طور که یوزپلنگ هنگام شکار بهترین طعمه را انتخاب می‌کند، این الگوریتم هم به طور هوشمند ویژگی‌هایی که بیشترین تأثیر را در تشخیص تقلب دارند شناسایی می‌کند و بقیه را حذف می‌کند.

برای مثال، فرض کنید ۱۰ ویژگی در تراکنش‌های کارت اعتباری وجود داشته باشد؛ این الگوریتم به صورت خودکار تصمیم می‌گیرد که مثلاً ۴ ویژگی از این ۱۰ مورد بیشتر از همه در تشخیص تقلب مؤثر هستند و باقی ویژگی‌ها را نادیده می‌گیرد. این باعث می‌شود که مدل به طور مؤثرتری آموزش ببیند و بتواند سریع‌تر و با دقت بیشتر کار کند.

۳) فاز طبقه‌بندی: در این فاز، داده‌ها با ویژگی‌های انتخابی به ماشین بردار پشتیبان جهت طبقه‌بندی داده می‌شود.

پس از انتخاب ویژگی‌های مهم، نوبت به طبقه‌بندی تراکنش‌ها می‌رسد. در این مرحله از ماشین بردار پشتیبان (SVM) استفاده می‌شود. این ابزار نوعی الگوریتم یادگیری ماشین است که هدف آن این است که تراکنش‌ها را به دو دسته تقسیم کند:

- تراکنش‌های قانونی
- تراکنش‌های تقلبی

ماشین بردار پشتیبان با استفاده از داده‌های آموزشی یاد می‌گیرد که چطور بین این دو دسته تفاوت قائل شود. سپس، وقتی یک تراکنش جدید دریافت می‌شود، این ماشین بردار پشتیبان تصمیم می‌گیرد که آیا این تراکنش قانونی است یا تقلبی.

نکته جالب در روش پیشنهادی شما این است که الگوریتم یوزپلنگ نه‌تنها در انتخاب ویژگی‌ها کمک می‌کند، بلکه به ماشین بردار پشتیبان هم کمک می‌کند تا بهترین پارامترها را برای طبقه‌بندی پیدا کند. این یعنی سیستم می‌تواند با دقت بیشتری تراکنش‌ها را دسته‌بندی کند.

برای انتخاب ویژگی برای هر جواب ممکن (عامل جستجو یا یوزپلنگ) یک بردار در نظر گرفته می‌شود که هر خانه آن مترادف با شماره ویژگی در مجموعه داده‌هاست:

ویژگی ۱	ویژگی ۲	ویژگی ۳	ویژگی ۴	ویژگی ۵	ویژگی ۶	ویژگی ۷	ویژگی ۸	ویژگی ۹	ویژگی ۱۰
۱	۰	۱	۱	۰	۰	۰	۱	۱	۰

شکل ۲. ساختار یوزپلنگ (جواب ممکن) در روش پیشنهادی در فاز انتخاب ویژگی

در شکل ۲ برای مجموعه داده‌ای با ۱۰ ویژگی، یک جواب ممکن یا یک یوزپلنگ یک آرایه با ۱۰ خانه است. در شکل ۱ یوزپلنگ به ویژگی‌های انتخابی اول، سوم، چهارم، هشتم و نهم اشاره دارد و معادل عدد باینری ۱۰۱۱۰۰۰۱۱۰ عدد حقیقی ۷۱۰ است که اپراتورهای الگوریتم یوزپلنگ وارد می‌شود. در روش پیشنهادی هر خانه از یوزپلنگ (جواب ممکن) که مقدار ۱ داشته باشد بدین معنی است که آن ویژگی در مجموعه ویژگی انتخاب قرار دارد. در روش پیشنهادی برای ارزیابی هر یوزپلنگ (مجموعه ویژگی) باید میزان برازندگی یا همان تابع برازندگی تخصیص یابد که برای این کار از روش نزدیک‌ترین همسایه برای ارزیابی مجموعه ویژگی‌های انتخابی با استفاده از معادله (۱) محاسبه می‌شود:

$$\text{fitness}_i(\text{selectedFeatures}) = \propto \left(\text{Accuracy}(\text{selectedFeatures})\right) + (1-\propto)\left(\frac{N_t - N_s}{N_t}\right) \qquad (۱)$$

که در آن N_t و N_s به ترتیب نماینده تعداد کل ویژگی‌ها و تعداد ویژگی‌های انتخاب‌شده و α یک ضریب بین ۰ و ۱ می‌باشند.

میزان Accuracy آمده در تابع برازندگی از رابطه (۲) به دست می‌آید:

$$\text{classificationAccuracy} = \frac{TN+\text{TP}}{\text{TN+FN+TP+FP}} \qquad (۲)$$

هر یک از عناصر ماتریس به شرح ذیل می‌باشد:

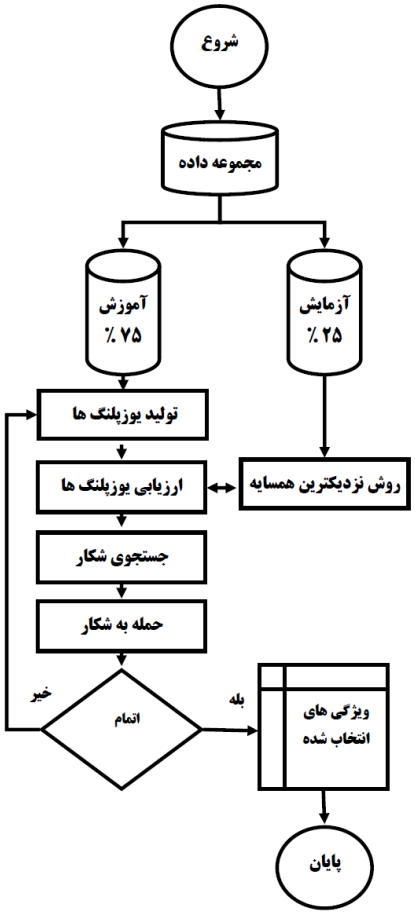
TN: بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به‌درستی منفی تشخیص داده است.

TP: بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به‌درستی مثبت تشخیص داده است.

FP: بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم دسته‌بندی دسته آن‌ها را به اشتباه مثبت تشخیص داده است.

FN: بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم دسته‌بندی دسته آن‌ها را به اشتباه منفی تشخیص داده است.

مراحل کلی روش پیشنهادی در انتخاب ویژگی در فلوچارت شکل ۳ آمده است:



شکل ۳. مراحل کلی روش پیشنهادی در انتخاب ویژگی با الگوریتم یوزپلنگ

در ماژول طبقه‌بندی (SVM-CO) ماشین بردار پشتیبان آموزش داده می‌شود. داده‌ها به بخش‌های آموزشی و آزمایشی تقسیم می‌شوند. بهبود روش ماشین بردار پشتیبان با استفاده از الگوریتم یوزپلنگ به این صورت است که در الگوریتم یوزپلنگ هر یوزپلنگ (جواب ممکن) یک مقدار تصادفی برای متغیر C و w در روش ماشین بردار پشتیبان در معادله ۱ تولید می‌شود، پارامتر C تنظیم‌کننده حاشیه هست که وظیفه آن برقراری تعادل بین حداکثر کردن حاشیه و حداقل کردن خطای دسته‌بندی بوده و همواره بزرگ‌تر از صفر است و پارامتر w هم‌وزن است و به علت آنکه در روش ماشین بردار پشتیبان به‌صورت تصادفی تولید می‌شود و ممکن است در بهترین مقدار خود قرار نگیرد، ازاین‌رو به دست آوردن مقدار مناسب C و w در روش ماشین بردار پشتیبان یک مسئله بهینه‌سازی است که الگوریتم یوزپلنگ بهترین مقدار را برای آن به دست می‌آورد. در ماشین بردار پشتیبان پیدا کردن بهترین w و C با کمینه‌سازی معادله ۳ محقق می‌شود:

$$\min \frac{1}{2}\|w\|^2 + C\sum_i \varepsilon_i \qquad (۳)$$

که در بهینه‌سازی آن باید شرط معادله ۴ در نظر گرفته شود:

$$y_i(<w.x_i>+b)\geq 1-\varepsilon_i \ , \ \varepsilon_i\geq 0 \quad \forall i \qquad (۴)$$

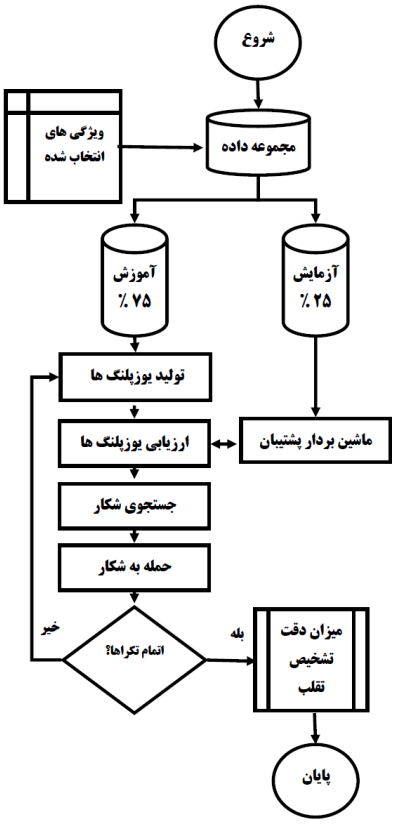
در معادله ۴ پارامتر b بایاس، x_i ویژگی داده و y_i کلاس داده است. در الگوریتم یوزپلنگ با استفاده از اپراتورهای خود، هر جواب ممکن (مقداری برای پارامتر C وw) را می‌یابد تا درنهایت به بهترین مقدار این پارامترها دست پیدا شود. برای محاسبه برازندگی هر جواب در این الگوریتم از تابع برازندگی میزان صحت طبقه‌بندی ماشین بردار پشتیبان از معادله ۲ استفاده می‌شود.

در مسئله تنظیم پارامترهای ماشین بردار پشتیبان، هر جواب ممکن در الگوریتم بهینه‌سازی یک ارایه است به‌صورت حقیقی که نشان‌دهنده مقدار عددی برای دو پارامتر W و C در ماشین بردار پشتیبان است:

C	W
0.45	0.74

شکل ۴ . ساختار یک یوزپلنگ در تنظیم پارامترهای ماشین بردار پشتیبان

در شکل ۴ هر یوزپلنگ نشان‌دهنده دو پارامتر اصلی در ماشین بردار پشتیبان است. مراحل کلی روش پیشنهادی در بهبود طبقه‌بندی ماشین بردار پشتیبان با الگوریتم بهینه‌سازی یوزپلنگ در فلوچارت شکل ۵ آمده است:



شکل ۵. مراحل تنظیم پارامتر ماشین بردار پشتیبان با الگوریتم یوزپلنگ

حرکات یوزپلنگ در الگوریتم بهینه‌سازی یوزپلنگ که در روش پیشنهادی برای انتخاب ویژگی و تنظیم پارامترهای ماشین بردار پشتیبان استفاده می‌شود، مشترک بوده و شامل:

- جستجوکردن: یوزپلنگ‌ها برای یافتن طعمه خود نیاز به جستجو، از جمله اسکن یا جستجوی فعال، در قلمرو خود (فضای جستجو) یا اطراف آن دارند.
- نشستن و انتظار: پس از شناسایی طعمه، اما وضعیت مناسب، یوزپلنگ‌ها ممکن است بنشینند و منتظر نزدیک شدن طعمه یا بهتر شدن وضعیت باشند.
- هجوم بردن: این استراتژی دو مرحله اساسی دارد:
  - عجله: زمانی که یوزپلنگ تصمیم به حمله می‌گیرد، با حداکثر سرعت به سمت طعمه می‌شتابد.
  - گرفتن: یوزپلنگ از سرعت و انعطاف‌پذیری برای گرفتن طعمه با نزدیک شدن به طعمه استفاده می‌کند.
- شکار را رها کند و به خانه برگردد: برای این استراتژی دو حالت در نظر گرفته شده است. (۱) اگر یوزپلنگ در شکار طعمه ناموفق باشد، باید موقعیت خود را تغییر دهد یا به قلمرو خود بازگردد. (۲) در مواردی که شکار موفقی در یک بازه زمانی انجام نشود، موقعیت خود را به آخرین شکار کشف شده می‌برد و جستجو را در اطراف آن انجام دهد.

استراتژی جستجو: یوزپلنگ‌ها از دو طریق به دنبال طعمه می‌گردند. یا در حالت نشسته یا ایستاده محیط را پایش کرده و یا به طور فعال در اطراف آن گشت زنی می‌کنند. حالت پایش زمانی مناسب‌تر است که طعمه هنگام راه رفتن در دشت متراکم باشد و در حال چرا باشد. از طرفی انتخاب حالت فعال که نیاز به انرژی بیشتری نسبت به حالت پایش دارد در صورتی که طعمه پراکنده و فعال باشد بهتر است. بنابراین، در طول دوره شکار، بسته به وضعیت طعمه، پوشش منطقه و وضعیت خود یوزپلنگ‌ها، زنجیره‌ای از این دو حالت جستجو است. معادله جستجوی تصادفی برای به‌روزرسانی موقعیت جدید یوزپلنگ در معادله ۶ آمده است که موقعیت فعلی با گام حرکتی بروز می‌شود:

$$X_{i,j}^{t+1} = X_{i,j}^{t} + \hat{r}_{i,j}^{-1}.\alpha_{i,j}^{t} \qquad (۶)$$

در معادله ۴ $X_{i,j}^{t+1}$ موقعیت بعدی یوزپلنگ و $X_{i,j}^{t}$ موقعیت فعلی آن است و $\hat{r}_{i,j}^{-1}$ پارامتر تصادفی با توزیع نرمال استاندارد است و $\alpha_{i,j}^{t}$ طول گام جهت حرکت است و بیشتر از ۰ است و حالت پیش‌فرض آن $\frac{t}{T} \times 0.001$ است به معنی اینکه یوزپلنگ در حال جستجوی آهسته است. همچنین ممکن است در مواجهه با شکارها و یا دشمنان دیگر، به سرعت حرکت کرده و تغییر جهت حرکت داشته باشد. $\alpha_{i,j}^{t}$ حرکتی بین یوزپلنگ و دیگر همسایه‌ها و یا رهبر است. رهبر به بهترین جواب پیدا شده در هر تکرار بهینه‌سازی گفته می‌شود.

استراتژی نشستن و منتظر ماندن: در طول حالت جستجو، طعمه ممکن است در میدان دید یوزپلنگ قرار گیرد. در این شرایط هر حرکت یوزپلنگ ممکن است طعمه را از حضور خود آگاه کند و منجر به فرار طعمه شود. برای جلوگیری از این نگرانی، یوزپلنگ ممکن است تصمیم بگیرد (با دراز کشیدن روی زمین یا پنهان شدن در میان بوته‌ها) کمین کند تا به اندازه کافی به طعمه نزدیک شود. بنابراین، در این حالت، یوزپلنگ در موقعیت خود باقی می‌ماند و منتظر می‌شود تا طعمه نزدیک‌تر شود، معادله ۷ برای این منظور در نظر گرفته شده است:

$$X_{i,j}^{t+1} = X_{i,j}^{t} \quad (۷)$$

در معادله ۷ $X_{i,j}^{t+1}$ موقعیت بعدی یوزپلنگ و $X_{i,j}^{t}$ موقعیت فعلی آن است و در واقع به‌روزرسانی در موقعیت یوزپلنگ‌ها رخ نمی‌دهد.

استراتژی حمله: یوزپلنگ‌ها از دو عامل مهم برای حمله به طعمه خود استفاده می‌کنند: سرعت و انعطاف‌پذیری. وقتی یوزپلنگ تصمیم به حمله می‌گیرد، با سرعت تمام به سمت طعمه می‌رود. پس از مدتی طعمه متوجه حمله یوزپلنگ می‌شود و شروع به فرار می‌کند. به عبارت دیگر، یوزپلنگ موقعیت شکار را دنبال می‌کند و جهت حرکت خود را به گونه‌ای تنظیم می‌کند که در یک نقطه راه شکار را مسدود کند. از آنجایی که یوزپلنگ با حداکثر سرعت به فاصله کمی از طعمه رسیده است، طعمه باید فرار کند و موقعیت خود را به طور ناگهانی تغییر دهد تا زنده بماند. یعنی موقعیت بعدی یوزپلنگ نزدیک آخرین موقعیت شکار است.

در معادله ۸ استراتژی حمله آمده است:

$$X_{i,j}^{t+1} = X_{B,j}^{t} + \check{r}_{i,j}.\beta_{i,j}^{t} \quad (۸)$$

که در آن $X_{B,j}^{t}$ موقعیت فعلی شکار است و در واقع بهترین موقعیت فعلی در الگوریتم است.$\check{r}_{i,j}$ عامل چرخش و $\beta_{i,j}^{t}$ بر عامل همکنش یوزپلنگ است. $X_{B,j}^{t}$ برای نزدیک شدن به طعمه که در واقع بهترین جواب مسئله است در نظر گرفته شد و $\beta_{i,j}^{t}$ نشان‌دهنده تعامل یوزپلنگ‌ها با دیگر یوزپلنگ‌ها و یا رهبر است. $\check{r}_{i,j}$ عامل چرخش نیز یک حرکت تصادفی با معادله ۹ است که در آن $r_{i,j}$ توزیع نرمال استاندارد است.

$$\check{r}_{i,j} = |r_{i,j}|^{\exp\left(\frac{r_{i,j}}{2}\right)} sin(2\pi r_{i,j}) \quad (۹)$$

در این الگوریتم برای حرکت‌های تصادفی از پارامترهای تصادفی r استفاده می‌شود و همچنین مقدار H با معادله ۱۰ استفاده می‌شود که در آن r_1 یک عدد تصادفی یکنواخت بین $[0و1]$ است.

$$H = e^{2\left(1-\frac{t}{T}\right)}(2r_1 - 1) \quad (۱۰)$$

شبه کد روش پیشنهادی در انتخاب ویژگی و تنظیم پارامترهای ماشین بردار پشتیبان به‌صورت زیر است:

۱) مشخص کردن ابعاد مسئله (در انتخاب ویژگی ابعاد به تعداد ویژگی‌های مجموعه داده) (در بهبود ماشین بردار پشتیبان ابعاد به تعداد ۲ بعد برای دو پارامتر W و C) و تعیین تعداد جمعیت اولیه یوزپلنگ‌ها

۲) تولید جمعیت اولیه به‌صورت تصادفی (برای انتخاب ویژگی به‌صورت باینری و برای بهبود ماشین بردار پشتیبان به‌صورت حقیقی بین ۰ و ۱) ارزیابی هر یوزپلنگ با تابع برازندگی (معادله ۱ برای انتخاب ویژگی و یا معادله ۲ برای بهبود ماشین بردار پشتیبان)

۳) بیشترین تکرارهای الگوریتم تعیین شود و مشخص کردن مقدار T و تا وقتی که به تکرار نهایی نرسیده است مراحل زیر انجام شود:

۱) مشخص کردن یوزپلنگ‌ها، رهبر آنها و طعمه با توجه به برازندگی آنها

۲) انتخاب تعداد تصادفی یوزپلنگ‌ها و برای هر یوزپلنگ مراحل زیر انجام شود (فقط در انتخاب ویژگی تبدیل عامل‌های جستجوی باینری به معادل حقیقی)

۱- مشخص کردن همسایه‌های هر یوزپلنگ

۲- محاسبه $\alpha, \beta, \hat{r}, \check{r}$ و H

۳- r_2, r_3 به‌صورت تولید تصادفی توزیع غیریکنواخت بین ۰ و ۱

۴- اگر $r_2 \leq r_3$

۵- تولید تصادفی توزیع غیریکنواخت بین ۰ و ۳ برای r_4 و اگر $H \geq r_4$

۶- انجام جستجو توسط هر یوزپلنگ با معادله ۶ در غیر این صورت انجام حرکت حمله به سمت شکار با معادله ۸

۷- اگر $H < r_4$ منتظر ماندن و حرکت نکردن با معادله ۷

۳) به‌روزرسانی رهبر

۴) شماره t یک عدد اضافه شود و اگر $t > rand \times T$ آنگاه رها کردن شکار

۴) اتمام الگوریتم و رهبر به‌عنوان بهترین عامل جستجو و جواب مسئله (در انتخاب ویژگی نشان‌دهنده ویژگی‌های انتخابی است)(در بهبود ماشین بردار پشتیبان بهترین مقادیر برای دو پارامتر W و C است)

۴ نتایج

مجموعه داده اعتبار استرالیا[1] که از مخزن یادگیری ماشین UCI برای این مطالعه تهیه شده است، شامل ۶۹۰ برنامه اعتباری است که هر کدام با ۱۵ ویژگی مشخص می‌شوند [49]. در میان اینها، شش ویژگی عددی، هشت مورد طبقه‌بندی هستند، و یک برچسب کلاس باینری نتیجه درخواست اعتبار را نشان می‌دهد (۱ برای تایید و ۰ برای رد)، همانطور که در جدول ۳ نشان داده شده است. ۳۰۷ نمونه درخواست‌های تایید شده و ۳۸۳ نمونه رد درخواست‌ها بودند که در این مجموعه داده ۶۹۰ نمونه را تشکیل می‌دهند. مجموعه داده‌ها به مجموعه داده آموزش با ۷۵ درصد نمونه‌ها با تعداد ۵۱۸ نمونه و آزمایش با ۲۵ درصد نمونه‌ها با تعداد ۱۷۲ تقسیم شد و از روش اعتبارسنجی متقاطع[2] ده برابری برای به گزارش نتیجه دقت تشخیص به‌صورت میانگین استفاده شد.

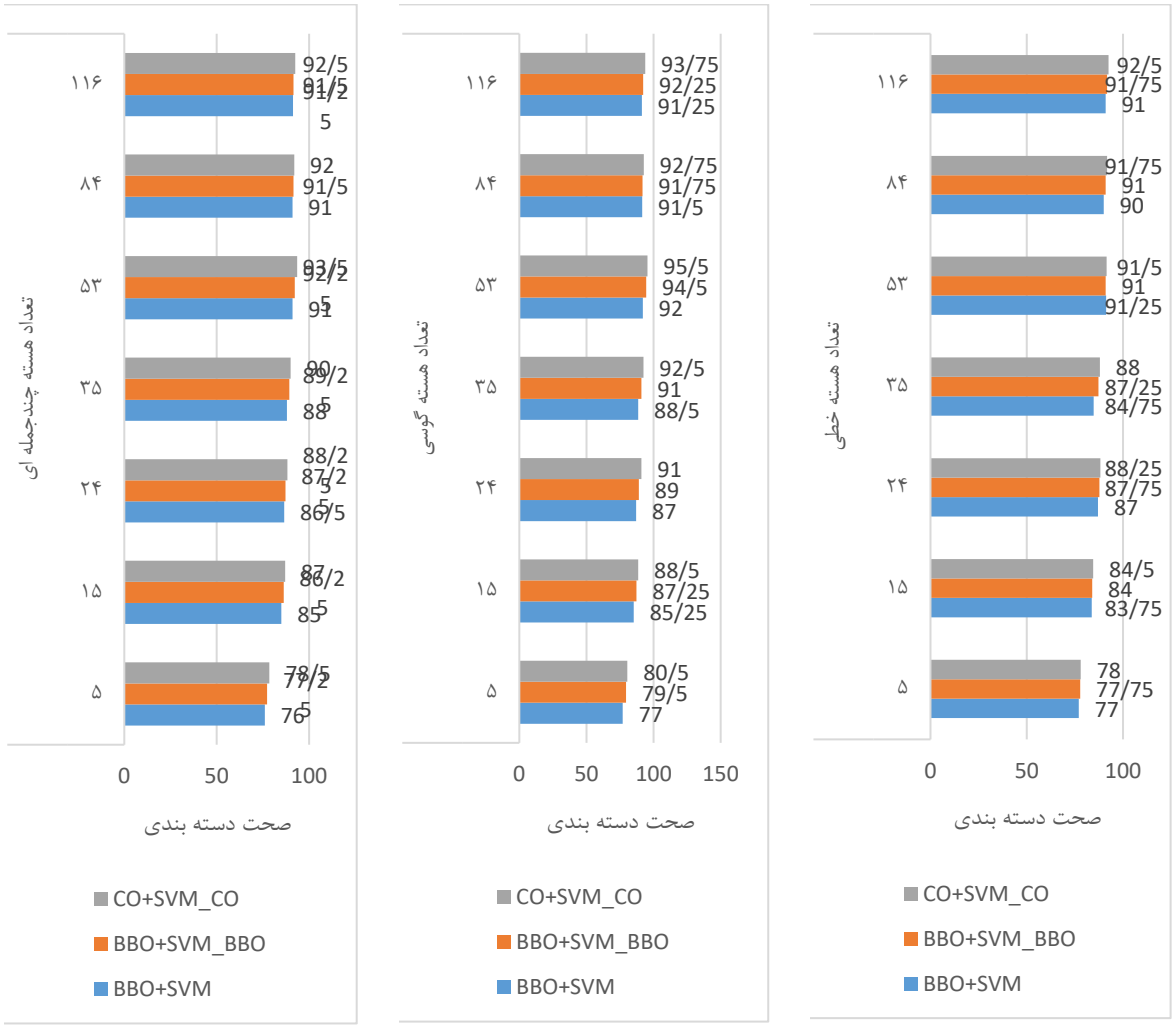
نتایج صحت تشخیص تقلب با معادله ۲ برای سه روش آمده است، که شامل:

- آزمایش با روش انتخاب ویژگی مبتنی بر الگوریتم خرس قهوه‌ای و طبقه‌بندی ماشین بردار پشتیبان بدون بهبود (روش مقاله [48]) با انتخاب هسته‌های مختلف گوسی، چندجمله‌ای و خطی
- آزمایش با روش انتخاب ویژگی مبتنی بر الگوریتم خرس قهوه‌ای و طبقه‌بندی ماشین بردار پشتیبان بهبود یافته با الگوریتم خرس قهوه‌ای با انتخاب هسته‌های مختلف گوسی، چندجمله‌ای و خطی

[1] Australian credit

[2] 10 fold cross validation

- آزمایش با روش انتخاب ویژگی مبتنی بر الگوریتم بهینه‌سازی یوزپلنگ و طبقه‌بندی ماشین بردار پشتیبان بهبود یافته با بهینه‌سازی یوزپلنگ (روش پیشنهادی) با انتخاب هسته‌های مختلف گوسی، چندجمله‌ای و خطی در شکل ۶ نتایج سه روش مبتنی بر ماشین بردار پشتیبان با تغییر تعداد هسته‌های مختلف آمده است.

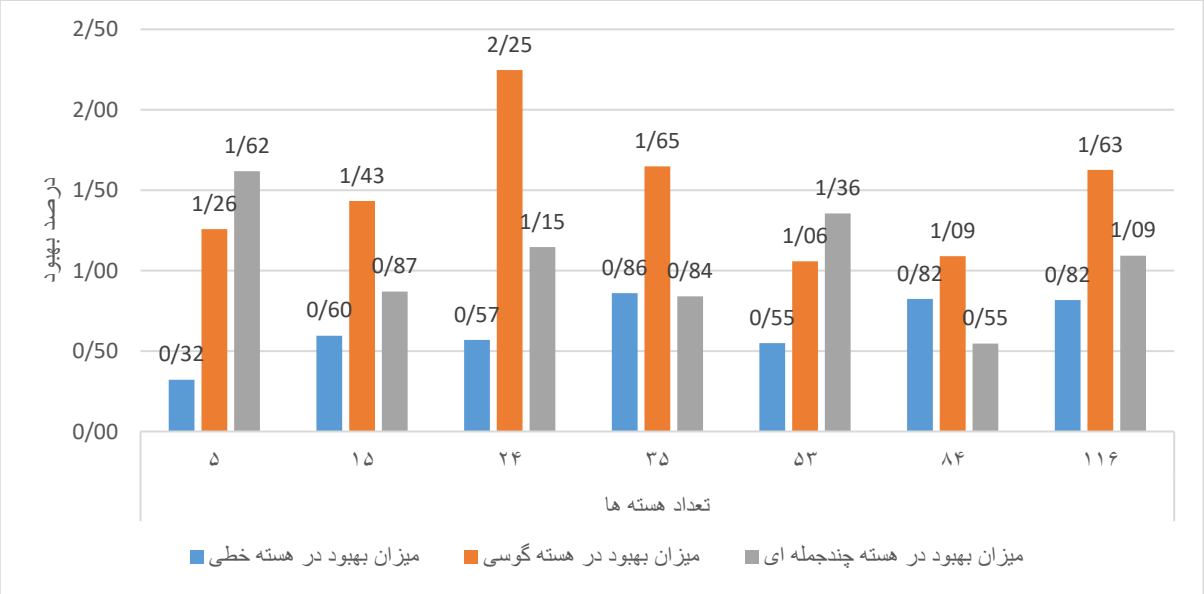


شکل ۶. نتایج صحت دسته‌بندی در هسته‌های مختلف ماشین بردار پشتیبان

BBO+SVM: انتخاب ویژگی با الگوریتم خرس قهوه‌ای و طبقه‌بندی با ماشین بردار پشتیبان بدون بهبود [۴۸]، BBO+SVM_BBO: انتخاب ویژگی با الگوریتم خرس قهوه‌ای و طبقه‌بندی با ماشین بردار پشتیبان بهبودیافته با الگوریتم خرس قهوه‌ای، CO+SVM_CO: انتخاب ویژگی با الگوریتم بهینه‌سازی یوزپلنگ و طبقه‌بندی با ماشین بردار پشتیبان بهبودیافته با الگوریتم بهینه‌سازی یوزپلنگ (روش پیشنهادی)

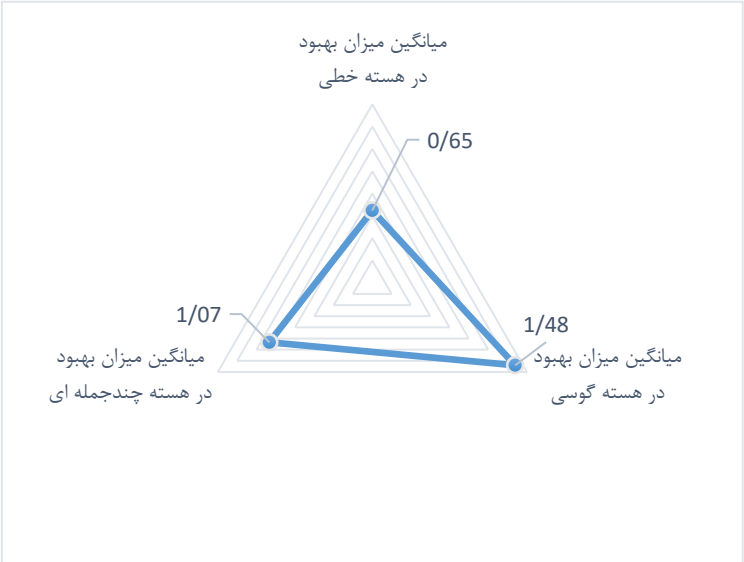
همانطور که از شکل ۶ مشخص است بیشترین مقدار صحت دسته‌بندی مربوط به روش CO+SVM_CO و سپس روش BBO+SVM_BBO و در آخر روش BBO+SVM [۴۸] است. در شکل ۶ میزان بهبود نتایج در

روش پیشنهادی و روش BBO+SVM_BBOآمده است. بالاترین میزان صحت نتایج عدد ۹۵/۵ درصد با هسته گوسی در روش ماشین بردار پشتیبان بهبودیافته با روش یوزپلنگ است.



شکل ۷. میزان بهبود نتایج در مقایسه با روش پیشنهادی و روش BBO+SVM_BBO

همانطورکه از شکل ۷ مشخص است بیشترین میزان بهبود نتایج ۲/۲۵ درصد و کمترین آن ۰/۳۲ است. به‌طور میانگین در آزمایشات انجام شده میزان بهبود نتایج روش پیشنهادی (CO+SVM_C) و روش BBO+SVM_BBO در شکل ۸ آمده است.



شکل ۸. میانگین میزان بهبود نتایج در مقایسه با روش پیشنهادی (CO+SVM_C) و روش BBO+SVM_BBO در هسته‌های مختلف

همانطور که از شکل ۸ مشخص است، بیشترین مقدار بهبود در هسته گوسی بوده است و در حالت کلی روش پیشنهادی توانسته در مجموعه آزمایشات نسبت به روشBBO+SVM_BBO که یک مرحله پیشرفته‌تر از روش BBO+SVM در مقاله [۴۸] است (در روش BBO+SVM_BBO بهبود ماشین بردار پشتیبان با الگوریتم خرس قهوه‌ای انجام شده است ولی در روش BBO+SVM در مقاله [۴۸] انجام نشده است)، نتایج بالاتری داشته باشد. نتایج نشان می‌دهد که ماشین بردار پشتیبان با هسته گوسی بهترین عملکرد را در روش پیشنهادی داشته است و توانسته به بهبود حداقل ۱/۴۸ درصدی نسبت به روش مقاله (BBO+SVM) [۴۸] و مدل پیشرفته آن یعنی روش BBO+SVM_BBO برسد، ولی در هسته خطی و چندجمله‌ای این میزان بهبود مشاهده نشده است. درواقع تأثیر الگوریتم بهینه‌سازی یوزپلنگ نسبت به الگوریتم بازپخت شبیه‌سازی شده در بهبود ماشین بردار پشتیبان وقتی‌که هسته گوسی انتخاب شده باشد، بیشتر است زیرا خط جداساز در هسته گوسی وابستگی بیشتری به تنظیم پارامتر W و C در ماشین بردار پشتیبان دارد ولی در هسته خطی این تأثیر کمتر است زیرا در هسته خطی، خط جداساز مانند هسته چندجمله‌ای و گوسی قدرت خمیدگی و جداسازی کلاس‌ها را ندارد.

۵ جمع‌بندی و نتیجه‌گیری

این مطالعه با استفاده از الگوریتم بهینه‌سازی یوزپلنگ یک راهکار جدید در انتخاب ویژگی و بهبود ماشین بردار پشتیبان ارائه شد. این رویکرد نتایج خوبی را به همراه داشت و قدرت تشخیص تقلب را به طور قابل‌توجهی افزایش داد زیرا توانست در انتخاب ویژگی مؤثرترین ویژگی‌ها و در طبقه‌بندی با ماشین بردار پشتیبان مقدار مناسب پارامترهای آن را پیدا کند. در این تحقیق مدل ارائه شده در تحقیق [۴۸] که در آن از انتخاب ویژگی با الگوریتم خرس قهوه‌ای و طبقه‌بندی با ماشین بردار پشتیبان بدون بهبود استفاده شده بود نیز در مدلی جدید ارتقا یافت که در آن بهبود ماشین بردار پشتیبان با الگوریتم خرس قهوه‌ای نیز انجام شد. نتایج نشان داده که روش پیشنهادی در مقایسه با دو روش دیگر نتایج دقت تشخیص تقلب بالاتری داشته است و بهترین نتیجه با هسته گوسی محقق شده است که در روش انتخاب ویژگی با الگوریتم خرس قهوه‌ای و تنظیم پارامترهای ماشین بردار پشتیبان با روش الگوریتم خرس قهوه‌ای در بهترین حالت ۹۴/۵ درصد صحت دسته‌بندی و روش پیشنهادی ۹۵/۵ درصد بوده است.

۶ پیشنهادهای کاربردی

به‌عنوان کارهای آینده می‌توان به استفاده از الگوریتم‌های فراابتکاری برای تنظیم تعداد هسته‌های ماشین بردار پشتیبان اشاره کرد. نتایج نشان می‌دهد که با تغییر هسته‌های ماشین بردار پشتیبان نتایج صحت دسته‌بندی تغییر می‌کند. بنابراین تعیین دقیق تعداد هسته یک مسئله بهینه‌سازی است که می‌تواند با الگوریتم‌های بهینه‌سازی بهبود یابد.

همچنین درصورتی‌که مجموعه داده استخراجی از تراکنش‌های بانکی ثبت گردد و همچنین نمونه‌های تقلب در آن مجموعه داده ثبت و یا شبیه‌سازی شود، مدل پیشنهادی می‌تواند بروی داده‌های بانک‌های داخلی آزمایش گردد. در حالت کلی مدل پیشنهادی قابلیت به‌کارگیری در داده‌کاوی برای تشخیص تقلب در انواع داده‌های جمع‌آوری شده مالی را دارد زیرا در فاز انتخاب ویژگی، ویژگی‌های مؤثر انتخاب شده و در فاز طبقه‌بندی الگوی داده‌های جدید آموزش می‌بیند تا بتواند به‌صورت عملی در تشخیص تقلب در سیستم بانکی استفاده شود.

فهرست منابع

[1] Song, Y., Escobar, O., Arzubiaga, U., & De Massis, A. (2022). The digital transformation of a traditional market into an entrepreneurial ecosystem. *Review of Managerial Science*, *16*(1), 65-88.

[2] Lucas, Y., & Jurgovsky, J. (2020). Credit card fraud detection using machine learning: A survey. arXiv preprint arXiv:2010.06479.

[3] Liu, Y., Gao, W., Hua, R., & Chen, H. (2021, March). Decomposition and measurement of economic effects of E-commerce based on static feder model and improved dynamic feder model. In *2021 2nd International Conference on E-Commerce and Internet Technology* (*ECIT*) (pp. 213-217). IEEE.

[4] Tran, L. T. T. (2021). Managing the effectiveness of e-commerce platforms in a pandemic. *Journal of Retailing and Consumer Services*, *58*, 102287.

[5] Pearlson, K. E., Saunders, C. S., & Galletta, D. F. (2024). *Managing and using information systems: A strategic approach*. John Wiley & Sons.

[6] Fanai, H., & Abbasimehr, H. (2023). A novel combined approach based on deep Autoencoder and deep classifiers for credit card fraud detection. *Expert Systems with Applications*, *217*, 119562.

[7] Singh, A., Jain, A., & Biable, S. E. (2022). Financial fraud detection approach based on firefly optimization algorithm and support vector machine. *Applied Computational Intelligence and Soft Computing*, *2022*(1), 1468015.

[8] Wahid, A., Msahli, M., Bifet, A., & Memmi, G. (2024). NFA: A neural factorization autoencoder based online telephony fraud detection. *Digital Communications and Networks*, *10*(1), 158-167.

[9] Carta, S., Fenu, G., Recupero, D. R., & Saia, R. (2019). Fraud detection for E-commerce transactions by employing a prudential Multiple Consensus model. *Journal of Information Security and Applications*, *46*, 13-22.

[10] Mutemi, A., & Bacao, F. (2024). E-Commerce Fraud Detection Based on Machine Learning Techniques: Systematic Literature Review. *Big Data Mining and Analytics*, *7*(2), 419-444.

[11] Asha, R. B., & KR, S. K. (2021). Credit card fraud detection using artificial neural network. *Global Transitions Proceedings*, *2*(1), 35-41.

[12] Yang, B., Gilles, H., & Bin, K. (2020). Artificial intelligence and fraud detection. *Innovative Technology at the Interface of Finance and Operations; Babich, V., Birge, J., Hilary, G., Eds*.

[13] Nandi, A. K., Randhawa, K. K., Chua, H. S., Seera, M., & Lim, C. P. (2022). Credit card fraud detection using a hierarchical behavior-knowledge space model. *Plos one*, *17*(1), e0260579.

[14] Agarwal, A., & Ratha, N. K. (2021, November). Black-Box Adversarial Entry in Finance through Credit Card Fraud Detection. In *CIKM Workshops*.

[15] Nayak, S., & Sharma, Y. K. (2023). A modified Bayesian boosting algorithm with weight-guided optimal feature selection for sentiment analysis. *Decision Analytics Journal*, *8*, 100289.

[16] Faris, H., Mafarja, M. M., Heidari, A. A., Aljarah, I., Ala'm, A. Z., Mirjalili, S., & Fujita, H. (2018). An efficient binary salp swarm algorithm with crossover scheme for feature selection problems. *Knowledge-Based Systems*, *154*, 43-67.

[17] Prakash, T., Singh, P. P., Singh, V. P., & Singh, S. N. (2023). A novel brown-bear optimization algorithm for solving economic dispatch problem. In *Advanced control & optimization paradigms for energy system operation and management* (pp. 137-164). River Publishers.

[18] Cartella, F., Anunciacao, O., Funabiki, Y., Yamaguchi, D., Akishita, T., & Elshocht, O. (2021). Adversarial attacks for tabular data: Application to fraud detection and imbalanced data. *arXiv preprint arXiv:2101.08030*.

[19] Beheshti, Z., & Shamsuddin, S. M. H. (2013). A review of population-based meta-heuristic algorithms. *Int. j. adv. soft comput. appl*, *5*(1), 1-35.

[20] P. Agrawal, H.F. Abutarboush, T. Ganesh, A.W. Mohamed, Metaheuristic algorithms on feature selection: A survey of one decade of research (2009–2019), IEEE Access 9 (2021) 26766–26791.

[21] Abualigah, L., Diabat, A., & Geem, Z. W. (2020). A comprehensive survey of the harmony search algorithm in clustering applications. *Applied Sciences*, *10*(11), 3827.

[22] Salcedo-Sanz, S. (2016). Modern meta-heuristics based on nonlinear physics processes: A review of models and design procedures. *Physics Reports*, *655*, 1-70.

[23] SamanehSorournejad, Z. Z., Atani, R. E., & Monadjemi, A. H. (2016). A survey of credit card fraud detection techniques: Data and technique oriented perspective.

[24] Akbari, M. A., Zare, M., Azizipanah-Abarghooee, R., Mirjalili, S., & Deriche, M. (2022). The cheetah optimizer: A nature-inspired metaheuristic algorithm for large-scale optimization problems. *Scientific reports*, *12*(1), 10953.

[25] Akshaya, B., & Sakthivel, P. (2024). VisionNet: An efficient vision transformer-based hybrid adaptive networks for eye cancer detection with enhanced cheetah optimizer. *Biomedical Signal Processing and Control*, *97*, 106673.

[26] Chen, S., Ji, Y., & Sun, X. (2024). Multi-User Detection Based on Improved Cheetah Optimization Algorithm. *Electronics*, *13*(10), 1842.

[27] Dimf, G. P., Kumar, P., & Joshua, K. P. (2023). CNN with BI-LSTM electricity theft detection based on modified cheetah optimization algorithm in deep learning. *SSRG Int J Electr Electron Eng*, *10*(2), 35-43.

[28] Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., & Anderla, A. (2019, March). Credit card fraud detection-machine learning methods. In *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)* (pp. 1-5). IEEE.

[29] Fang, Y., Zhang, Y., & Huang, C. (2024). Credit card fraud detection based on machine learning. In *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence* (p. 35). Springer Nature.

[30] Balogun, A. O., Basri, S., Abdulkadir, S. J., & Hashim, A. S. (2019). Performance analysis of feature selection methods in software defect prediction: a search method approach. *applied sciences*, *9*(13), 2764.

[31] Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M. S., & Zeineddine, H. (2019). An experimental study with imbalanced classification approaches for credit card fraud detection. *IEEE Access*, *7*, 93010-93022.

[32] Forough, J., & Momtazi, S. (2021). Ensemble of deep sequential models for credit card fraud detection. *Applied Soft Computing*, *99*, 106883.

[33] Khalilia, M., Chakraborty, S., & Popescu, M. (2011). Predicting disease risks from highly imbalanced data using random forest. *BMC medical informatics and decision making*, *11*, 1-13.

[34] Seera, M., Lim, C. P., Kumar, A., Dhamotharan, L., & Tan, K. H. (2024). An intelligent payment card fraud detection system. *Annals of operations research*, *334*(1), 445-467.

[35] DEB, K., Ghosal, S., & Bose, D. (2021). A comparative study on credit card fraud detection.

[36] Rtayli, N., & Enneya, N. (2020). Selection features and support vector machine for credit card risk identification. *Procedia Manufacturing*, *46*, 941-948.

[37] Ileberi, E., Sun, Y., & Wang, Z. (2022). A machine learning based credit card fraud detection using the GA algorithm for feature selection. *Journal of Big Data*, *9*(1), 24.

[38] Khan, M. Z., Shaikh, S. A., Shaikh, M. A., Khatri, K. K., Rauf, M. A., Kalhoro, A., & Adnan, M. (2022). The Performance Analysis of Machine Learning Algorithms for Credit Card Fraud Detection. *iJOE*, *19*(03), 83.

[39] Mahesh, K. P., Afrouz, S. A., & Areeckal, A. S. (2022). Detection of fraudulent credit card transactions: A comparative analysis of data sampling and classification techniques. In *Journal of Physics: Conference Series* (Vol. 2161, No. 1, p. 012072). IOP Publishing.

[40] Noviandy, T. R., Idroes, G. M., Maulana, A., Hardi, I., Ringga, E. S., & Idroes, R. (2023). Credit card fraud detection for contemporary financial management using xgboost-driven machine learning and data augmentation techniques. *Indatu Journal of Management and Accounting*, *1*(1), 29-35.

[41] Agarwal, S. (2023). An Intelligent Machine Learning Approach for Fraud Detection in Medical Claim Insurance: A Comprehensive Study. *Scholars Journal of Engineering and Technology*, *11*(9), 191-200.

[42] Yi, Z., Cao, X., Pu, X., Wu, Y., Chen, Z., Khan, A. T., ... & Li, S. (2023). Fraud detection in capital markets: A novel machine learning approach. *Expert Systems with Applications*, *231*, 120760.

[43] Valavan, M., & Rita, S. (2023). Predictive-Analysis-based Machine Learning Model for Fraud Detection with Boosting Classifiers. *Computer Systems Science & Engineering*, *45*(1).

[44] Khalid, A. R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J., & Adejoh, J. (2024). Enhancing credit card fraud detection: an ensemble machine learning approach. *Big Data and Cognitive Computing*, *8*(1), 6.

[45] Cherif, A., Ammar, H., Kalkatawi, M., Alshehri, S., & Imine, A. (2024). Encoder–decoder graph neural network for credit card fraud detection. *Journal of King Saud University-Computer and Information Sciences*, *36*(3), 102003.

[46] Mim, M. A., Majadi, N., & Mazumder, P. (2024). A soft voting ensemble learning approach for credit card fraud detection. *Heliyon*, *10*(3).

[47] Sinap, V. (2024). Comparative analysis of machine learning techniques for credit card fraud detection: Dealing with imbalanced datasets. *Turkish Journal of Engineering*, *8*(2), 196-208.

[48] Sorour, S. E., AlBarrak, K. M., Abohany, A. A., & Abd El-Mageed, A. A. (2024). Credit card fraud detection using the brown bear optimization algorithm. *Alexandria Engineering Journal*, *104*, 171-192.

[49] Asuncion, A., & Newman, D. (2007, November). *UCI machine learning repository*.

Improving Machine Learning in Credit Card Fraud Detection

Amin Aref[1] **Farshid Farahkhah**[2]

Abstract

Fraud detection in banking systems is critical to financial stability, customer protection, and regulatory compliance. Machine learning is critical in improving data analytics as well as real-time fraud detection. Feature selection is an important phase in machine learning to enhance credit card fraud detection. In fact, eliminating the negative effect of additional and irrelevant features and selecting effective features in the feature selection phase helps the classification phase in machine learning. In this article, an effective method based on the cheetah optimization algorithm is presented to increase the capacity to accurately identify credit card fraud transactions by recognizing the relevant features. Also, in this research, the support vector machine method is in the classification phase of machine learning, which is improved by adjusting the parameters using the cheetah algorithm. The results have shown that the proposed method on the Australian credit data set has at least 1.48% improvement in fraud detection compared to other methods.

Keywords: Credit Card Fraud Detection, Machine Learning, Cheetah Optimization Algorithm
JEL Classification: G21, E58, E51

[1] A.aref@shaparak.com

[2] f.farahkhah@shaparak.com