دوفصلنامه محاسبات و سامانه های توزیع شده
سال هشتم، شماره۲، شماره پیاپی۱۶، صفحه ۸۸ تا ۱۰۲، سال ۱۴۰۴

ارائه یک مدل ریسک اعتباری بانکی جهت طبقه‌بندی مشتریان مبتنی بر الگوریتم تکاملی کلاغ و شبکه عصبی کانولوشن

سحر کریم‌لو سیاح
کارشناسی ارشد رسته مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیکز از دانشگاه آزاد اسلامی واحد تهران جنوب

چکیده

ارزیابی ریسک اعتباری مشتریان حیاتی‌ترین و مهم‌ترین وظایف در صنعت بانکداری مدرن است که تضمین مطالبات غیرجاری و پایداری مالی مؤسسات اعتباری با یک پیش‌بینی دقیق از وضعیت اعتباری متقاضیان وام را به دنبال دارد. هدف اصلی مقاله، ارتقای فرآیند ارزیابی ریسک اعتباری برای شرکت‌های کوچک است که یک اقدام نوآورانه در این زمینه است و رای دستیابی به این هدف، یک مجموعه داده جامع شامل ابعاد مختلفی مانند اطلاعات مالی، اطلاعات خرد رفتاری شرکت‌ها، اطلاعات اعتبار عمومی و دسترسی شخص ثالث به اطلاعات لازم است تا بتوان از تکنیک‌های پردازش نمونه نامتوازن برای دستیابی به نمایش عادلانه‌تر نمونه‌های طبقه اقلیت استفاده کرد. در این مدل، یک رویکرد دومرحله‌ای مبتنی بر الگوریتم بهینه‌سازی فراابتکاری کلاغ و شبکه عصبی عمیق کانولوشن طراحی و پیاده‌سازی شد که در مرحله اول، از الگوریتم کلاغ برای حل مشکل انتخاب ویژگی استفاده شد و در مرحله دوم، ویژگی‌های منتخب به عنوان ورودی به یک شبکه عصبی کانولوشن برای طبقه‌بندی نهایی ارائه شد. مدل پیشنهادی بر روی مجموعه داده معتبر برگرفته از پلتفرم کگل[1] که شامل اطلاعات واقعی مشتریان بانکی بود، پیاده‌سازی و ارزیابی شد. برای سنجش دقیق عملکرد مدل از روش اعتبارسنجی متقابل K-Fold استفاده شد. نتایج آزمایش‌ها نشان داد که این مدل ترکیبی در مقایسه با روش‌های طبقه‌بندی سنتی، به طور معناداری به دقت بالاتری دست یافته است. در نهایت، این پژوهش نشان داد که مدل پیشنهادی، پتانسیل بالایی برای استفاده به عنوان یک سیستم پشتیبان تصمیم در بانک‌ها و مؤسسات مالی جهت ارزیابی ریسک اعتباری دارد.

واژه‌های کلیدی: پیش‌بینی و ارزیابی ریسک اعتباری، طبقه‌بندی مشتریان بانکی، الگوریتم تکاملی کلاغ و شبکه عصبی کانولوشن.

تاریخچه مقاله:
تاریخ ارسال: ۱۴۰۴/۰۹/۱۵
تاریخ اصلاحات: ۱۴۰۴/۱۰/۲۵
تاریخ پذیرش: ۱۴۰۴/۱۱/۱۵
تاریخ انتشار: ۱۴۰۴/۱۱/۲۰
ایمیل نویسنده مسئول:sayah.sahar@gmail.com

۱ - مقدمه و بیان مسئله

در واقع اصطلاح «اعتبار» در صنعت بانکداری و مالی به قراردادی اطلاق می‌شود که در آن یک سازمان مالی مقدار محدودی پول را در اختیار افراد یا سازمان‌ها قرار می‌دهد، با این تعهد که وجوه استقراضی بعداً به‌صورت یکجا یا از طریق چند اقساط بازپرداخت می‌شود. وام‌دهی مستلزم سطح معینی از ریسک است، زیرا مستلزم اتخاذ تصمیمات پیچیده است که به طور بالقوه می‌تواند منجر به ورشکستگی موسسه مالی شود. بنابراین، ارزیابی دقیق عملیات اعتباری بسیار مهم است. با توجه به این اهمیت، بانک‌ها محدودیت‌های اعتباری ایجاد می‌کنند و سازمان‌های اقتصادی باید عملیات اعتباری خود را با ظرفیت ریسک خود هماهنگ کنند. در نتیجه، متخلفان احتمالی را می‌توان بر اساس عواملی مانند سطح ریسک، وثیقه ارائه شده و ماهیت معاملات مالی انجام شده شناسایی کرد[۱]. مدیریت ریسک مالی موضوع حساسی است که باید مورد بررسی قرار گیرد. برخی از سازمان‌ها، صنایع و دولت‌ها در سراسر جهان به سیستم‌های مدیریت ریسک و امتیازدهی اعتباری وابسته هستند. کلاهبرداری مالی، که شامل کلاهبرداری تجاری، کلاهبرداری وام های شخصی، پولشویی، کلاهبرداری کارت اعتباری، تقلب بیمه، تقلب در وام همتا به همتا و موارد دیگر

[1] Kaggle

است، یک استراتژی آگاهانه، مقصر یا کلاهبرداری است که به قصد بهره برداری از ساختار یک سازمان غیرانتفاعی به منظور دستیابی به منافع مالی غیرقانونی و غیرقانونی انجام می‌شود. برای مدیریت ریسک‌های مالی و تصمیم‌گیری در مورد وام دادن پول، بانک‌ها و سایر موسسات مالی باید داده‌های مشتریان را جمع‌آوری کنند. این روش می‌تواند به شناسایی بدهکاران خوب و بد کمک کند. بانک‌ها مشتریان «قرض گیرنده خوب» را با سابقه اعتباری تمیز در نظر می‌گیرند و وام گیرندگان بد» اعتبار ضعیفی دارند[۲].

ارزیابی ریسک اعتباری مصرف‌کننده نقش مهمی در کنترل ریسک و مدیریت شرکت‌های مالی مصرف‌کننده، مؤسسات وام کوچک، بانک‌ها و سایر مؤسسات مالی (به عنوان مثال، امتیازدهی برنامه‌ها، و امتیازدهی رفتاری) ایفا می‌کند. ارزش اقتصادی این کار بسیار زیاد است، به عنوان مثال، کل مبلغ اعتبار مصرفی معوق تا اوت ۲۰۲۲ در ایالات متحده از ۴۷۰ میلیارد دلار فراتر رفت، که نسبت به سال ۲۰۲۱ بیش از ۶ درصد افزایش یافته است. در عین حال، نرخ معوقه وام های مصرفی توسط بانک های تجاری رشد بیش از ۱۳ درصدی را از سال ۲۰۲۱ توسط کمیسیون نظارت بر چین تجربه کرده است. نسبت وام بانک‌های تجاری چین در سه ماهه سوم سال ۲۰۲۲ به حدود ۱.۶۶ درصد رسید. موسسات مالی باید مدل‌های عالی را برای حمایت از تصمیم‌گیری و به حداقل رساندن زیان های اقتصادی احتمالی ناشی از افزایش نکول مصرف کننده و نسبت وام‌های بد ایجاد کنند. برای دستیابی به این هدف، از مدل‌های امتیازدهی اعتباری (یعنی کارت‌های امتیازی) برای تخمین احتمال نکول وام‌گیرنده در آینده با پیش‌بینی ریسک نکول آنها استفاده می‌شود[۳].

در دهه‌های اخیر، محققان زیادی بر روی این موضوع تمرکز کرده‌اند و تلاش می‌کنند تا راه‌حل مفیدی برای این مشکل ارائه کنند، الگوریتم‌های طبقه‌بندی مختلف در یادگیری ماشین به طور مداوم برای بهبود دقت پیش‌بینی توسعه می‌یابند[۳]. در واقع داده کاوی یک حوزه اصلی تحقیقاتی است که هدف آن جمع‌آوری و تجزیه و تحلیل مقادیر زیادی داده با یادگیری ماشین است. داده کاوی در حال حاضر در بخش بانکداری رایج است، زیرا تکنیک‌های تحلیلی موفقی برای شناسایی دانش مربوط به امتیازات اعتباری وجود دارد. هدف اولیه بهره‌برداری از حجم عظیمی از اطلاعات موجود در پایگاه های داده برای تصمیم گیری‌های کلیدی است[۴]. از آنجایی که داده‌ها پایه و اساس مدل‌سازی را تشکیل می‌دهند، عصر داده‌های بزرگ مجموعه‌های داده‌ای را با ویژگی‌های متنوع معرفی کرده است که دشواری مدل‌سازی را افزایش داده است. در نتیجه، ادبیات الگوریتم‌ها و ویژگی‌های داده‌ها ممکن است برای محققان در این زمینه آشفته به نظر برسد و آنها را دچار سردرگمی کند. بنابراین، بررسی متفکرانه مقالات مرتبط برای کمک به محققان در درک موضوع و توسعه ارزیابی ریسک اعتباری مصرف‌کننده ضروری است[۳]. در این پژوهش به بررسی جدیدترین تحقیقات از دیدگاه جدیدی می‌پردازد و عمدتاً بر روی برخی از مطالعات با پیش‌بینی اینکه آیا مصرف‌کنندگان در آینده در طول فرآیند وام‌گیری، که اغلب به عنوان یک مشکل طبقه‌بندی باینری در نظر گرفته می‌شود، نکول می‌کنند یا خیر، تمرکز دارد. یک مدل مناسب پیشنهاد می‌دهد که به دنبال پاسخ این سوال است که "آیا ارائه پیش‌بینی و ارزیابی ریسک اعتباری جهت طبقه‌بندی مشتریان بانکی مبتنی بر الگوریتم - تکاملی کلاغ و شبکه عصبی کانولوشن امکان پذیر است؟"

۲ - کارهای مرتبط

در مقاله‌ای به مقایسه الگوریتم برای طبقه بندی داده کاوی: ارزیابی ریسک پیش فرض امتیازدهی اعتباری مشتریان بانک پرداخته شد. نویسندگان مقاله معتقدند که رتبه بندی ریسک اعتباری مصرف کننده شامل ارزیابی ریسک های درخواست اعتبار است. بنابراین، هر کسب و کاری باید بدهکاران و غیر بدهکاران را به درستی شناسایی کند. این مقاله از رویکردهای یادگیری ماشین برای شبیه‌سازی ریسک اعتباری مصرف‌کننده استفاده می‌کند و نتایج را با مدل لجستیک مقایسه می‌کند، و تعیین می‌کند که آیا یادگیری ماشین رتبه‌بندی پیش‌فرض مشتری را بهبود می‌بخشد یا خیر. این مقاله به بررسی چگونگی تأثیر ویژگی‌های مشتری بر تجربیات مجازی می پردازد. با وجود پیشرفت در مدل‌های یادگیری ماشینی برای ارزیابی اعتبار، مجموعه داده‌های نامتعادل و شکست برخی الگوریتم‌ها در توضیح پیش‌بینی‌ها همچنان مسائل اصلی هستند. این مطالعه از تحصیلات، سن، وضعیت تأهل، سابقه پرداخت و جنسیت مصرف کنندگان کارت اعتباری تایوانی ۲۰۰۵ استفاده کرد. آزمایش پیش‌فرض با استفاده از رگرسیون لجستیک، k- نزدیکترین همسایگان، ماشین بردار پشتیبان،

درخت تصمیم، جنگل تصادفی، طبقه‌بندی کننده تقویت کننده آدابوست و تقویت گرادیان مدل‌سازی می‌شود. دقت، صحت، فراخوان، منحنی ویژگی عملیاتی گیرنده و منحنی دقت فراخوان مدل‌ها مورد ارزیابی قرار گرفت. رتبه بندی متریک منحنی ویژگی عملیاتی گیرنده جنگل تصادفی ۹۷٪ از تمام معیارهای دقت دیگر بهتر عمل کرد. مدل لجستیک عملکرد ضعیفی داشت، در حالی که یادگیری ماشینی طبقه‌بندی پیش فرض را بهبود بخشید[۲].

در مقاله‌ای، آخرین پیشرفت ها در ارزیابی ریسک اعتباری با یادگیری ماشین و تکنیک های یادگیری عمیق ارائه شد. نویسندگان مقاله معتقدند که وام برای افراد و سازمانها برای دستیابی به اهدافشان حیاتی است. با این حال، موسسات مالی با چالش‌هایی مانند مدیریت زیان‌ها و فرصت‌های از دست رفته در تصمیم‌گیری‌های وام مواجه هستند. یک مسئله کلیدی، مجموعه داده‌های نامتعادل در ارزیابی ریسک اعتباری است که مانع از پیش‌بینی دقیق افراد متخلف می‌شود. تحقیقات قبلی از تکنیک‌های یادگیری ماشینی، از جمله سیستم‌های طبقه‌بندی‌کننده منفرد یا چندگانه، روش‌های مجموعه، و رویکردهای متعادل‌سازی کلاس استفاده کرده‌اند. این بررسی عوامل مختلف و روش‌های یادگیری ماشینی را برای ارزیابی ریسک اعتباری خلاصه می‌کند که در قالب جدولی ارائه شده است تا بینش‌های ارزشمندی را برای محققان ارائه کند. پیچیدگی داده‌ها، توزیع کلاس اقلیت، تکنیک‌های نمونه‌گیری، انتخاب ویژگی و پارامترهای فرا یادگیری را پوشش می‌دهد. هدف کمک به توسعه الگوریتم‌های جدید است که از روش‌های موجود بهتر عمل می‌کنند. حتی یک بهبود جزئی در نرخ‌های پیش‌بینی نکول‌ها می‌تواند با صرفه‌جویی میلیون‌ها نفر برای وام‌دهندگان، جامعه را به‌طور چشمگیری تحت تأثیر قرار دهد[۱].

مقاله‌ای به یادگیری ماشین برای تجزیه و تحلیل ریسک اعتباری در سراسر ایالات متحده پرداخت. نویسندگان مقاله معتقدند که موسسات مالی وام دهنده با ریسک مواجه هستند که نیازمند اتخاذ یک چارچوب قوی برای مدیریت موثر آن است. در حالی که روش های سنتی در سراسر صنعت مالی به کار گرفته شده است، ظهور هوش مصنوعی این فرصت را به سازمان‌ها می‌دهد تا از روش‌های پیشرفته برای مدیریت ریسک اعتباری استفاده کنند. این مقاله بر روی کاربرد تکنیک‌های یادگیری ماشین برای تحلیل ریسک اعتباری تمرکز دارد. داده های ثانویه در مورد اطلاعات مربوط به وام گیرندگان از پایگاه داده کگل استخراج شد که داده های دفتر اعتباری را شبیه سازی می‌کند. دو مدل مجموعه در جنگل تصادفی و تقویت گرادیان برای این مطالعه اتخاذ شد. یافته‌ها نشان داد که درصد درآمد برای بازپرداخت وام، درآمد وام گیرنده و نرخ سود وام‌ها از مهم‌ترین ویژگی‌ها برای تعیین معوقان هستند. علاوه بر این، نتایج ارزیابی نشان داد که هر دو الگوریتم جنگل تصادفی و تقویت گرادیان، با نمرات F1 به ترتیب ۹۲.۹ درصد و ۹۳ درصد، عملکرد خوبی داشتند. توصیه شد که مؤسسات مالی باید تأیید و جامعیت داده‌های خود را در اولویت قرار دهند، زیرا داده‌های دقیق برای توسعه مدل‌های تاب‌آور ضروری است[۵].

مقاله‌ای به ارزیابی ریسک اعتباری مالی بر اساس یادگیری ماشینی پرداخت. نویسندگان مقاله معتقدند که عصر کلان داده، صنعت مالی با چالش‌ها و فرصت‌های جدیدی روبرو است. از طریق کلان داده و فناوری هوش مصنوعی، می‌توان انواع ریسک‌های مالی از جمله ریسک اعتباری، ریسک بازار، ریسک تقلب و غیره را با دقت بیشتری ارزیابی و مدیریت کرد. در این مقاله از مدل درخت تصمیم برای مدلسازی و تحلیل ریسک اعتباری در ریسک مالی استفاده شده و سیستم پیشگیری و کنترل ریسک مالی ایجاد شده است. برای ریسک اعتباری، با توجه به داده‌های اطلاعات مشتریان بانک، پس از پردازش داده‌ها، از روش طبقه‌بندی درخت تصمیم برای قضاوت در مورد اینکه آیا مشتری ممکن است در آینده از طریق اطلاعات اولیه مشتری نکول کند یا خیر، استفاده می‌شود و در نهایت مبنای اصلی تمایز سن، درآمد مورد انتظار، موجودی و تعداد کارت‌های اعتباری است. سپس از ابعاد پنج گانه ایجاد ارزیابی صحیح ریسک و سازوکار هشدار زودهنگام، ارتقای آگاهی شهروندان از ریسک مالی، ترویج ساخت صندوق های تضمین ثبات مالی، تقویت نظارت و خود انضباطی صنعت و بهبود نظام قانونی و نظارتی، پیشنهادات قابل اجرا برای ریسک های مالی ارائه شده است[۶].

در مقاله‌ای دیگر اندازه‌گیری ریسک اعتباری، تجزیه و تحلیل تصمیم‌گیری، تحول و ارتقای داده‌های بزرگ مالی ارائه شد. نویسندگان مقاله معتقدند که اندازه‌گیری ریسک اعتباری

و تحلیل تصمیم‌گیری برای کلان داده‌های مالی، تئوری خوبی وجود ندارد و یک سیستم ارزیابی مؤثر و علمی برای آن‌ها شکل نگرفته است و مرور آنها می‌تواند به درک موضوعات مذکور، درک مسائل جاری، تجزیه و تحلیل مسائل پژوهشی، تسلط بر چالش‌های پژوهشی و پیش‌بینی جهت‌گیری‌های تحقیقاتی آینده کمک کند. همچنین، این مقاله به چهار جهت تحقیقاتی اندازه‌گیری ریسک اعتباری و تحلیل تصمیم‌گیری برای داده‌های بزرگ مالی اشاره می‌کند و می‌تواند راهنمایی‌ها و بینش‌هایی را برای پزشکان، محققان، مؤسسات مالی و بخش‌های دولتی که به تصمیم‌گیری پیچیده در کلان داده علاقه دارند، ارائه دهد[۷].

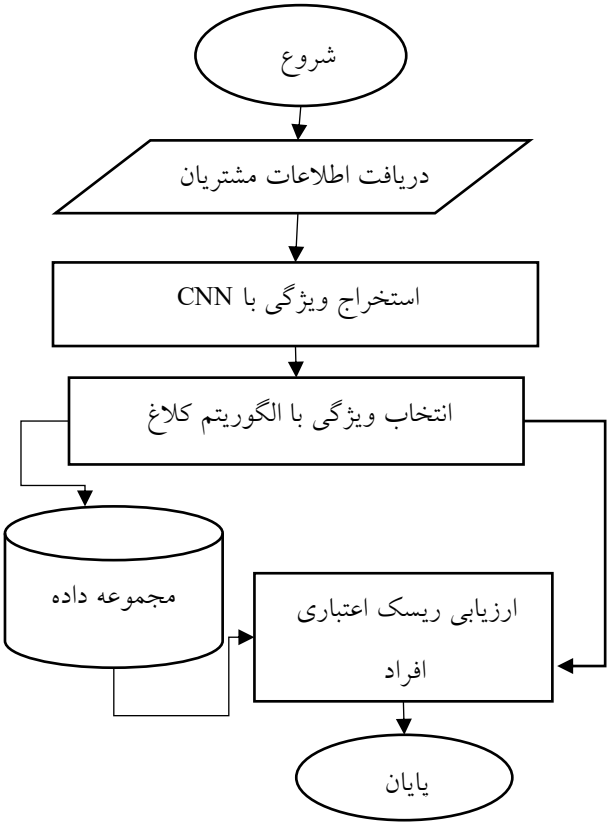
۳ - تحلیل روش پیشنهادی

شرکت‌های کوچک و خرد، محرک‌های حیاتی پیشرفت اقتصادی و ایجاد شغل در سطح جهانی هستند. طبق گزارش بانک جهانی، این کسب‌وکارها تقریباً ۹۰ درصد از کل کسب‌وکارهای جهان را تشکیل می‌دهند که بیش از نیمی از اشتغال جهانی را به خود اختصاص می‌دهند. بخش شرکت‌های کوچک و خرد چین در دو دهه گذشته رشد بی‌سابقه‌ای داشته است. امروزه، این شرکت‌ها ۹۵.۸۰ درصد از کل کسب‌وکارهای این کشور را تشکیل می‌دهند. با این وجود، علیرغم تأثیر قابل توجه آنها بر چشم‌انداز اقتصادی و چارچوب اجتماعی چین، این شرکت‌ها هنگام جستجوی حمایت مالی، به ویژه از بانک‌های تجاری، با موانع قابل توجهی روبرو می‌شوند. چنین چالش‌های تأمین مالی به طور مداوم مانع از پتانسیل رشد آنها می‌شود. رابطه روشنی بین این مشکلات تأمین مالی، هزینه‌های بالای تأمین مالی و ارزیابی‌های ریسک اعتباری که برای آنها در نظر گرفته شده است، پدیدار می‌شود. با کاهش ریسک اعتباری ادراک شده برای یک شرکت کوچک یا خرد، هزینه‌های تأمین مالی که متحمل می‌شوند نیز کاهش می‌یابد و راه‌های آنها را برای تأمین بودجه گسترش می‌دهد[۷].

ادغام کلان‌داده، یادگیری ماشین و داده‌کاوی یک سه‌گانه پویا ایجاد می‌کند که نوآوری را پیش می‌برد، کارایی را افزایش می‌دهد و تصمیم‌گیری را در دنیای داده‌محور ما بهبود می‌بخشد. بهره‌گیری از قدرت این فناوری‌ها، اثرات دگرگون‌کننده‌ای با تأثیرات گسترده در صنایع و زمینه‌های تحقیقاتی مختلف دارد[۸]. کلان‌داده که با حجم، سرعت و تنوع خود مشخص می‌شود، مخزن عظیمی برای استخراج اطلاعات فراهم می‌کند که به ایجاد زمینه برای فناوری‌های دگرگون‌کننده کمک کرده است. توانایی تجزیه و تحلیل و استفاده از مجموعه داده‌های پیچیده، که در دنیای به هم پیوسته امروز به صورت تصاعدی رشد می‌کنند، به سازمان‌ها و صنایع فرصت‌های بی‌نظیری برای کسب بینش‌های ارزشمند، تصمیم‌گیری‌های داده‌محور و کشف الگوهای ناشناخته قبلی ارائه می‌دهد. از تجزیه و تحلیل پیش‌بینی‌کننده گرفته تا پردازش بلادرنگ، تجزیه و تحلیل کلان‌داده به ابزاری ضروری برای پیشبرد نوآوری و کارایی در بخش‌های مختلف تبدیل شده است. همچنین به آماده‌سازی داده‌ها، درک کسب‌وکار، مدل‌سازی داده‌ها و تجزیه و تحلیل اکتشافی برای استخراج بینش‌های پنهان و شناسایی روندها برای استخراج اطلاعات معنادار کمک می‌کند. کشف گنجینه دانش درون داده‌ها، استفاده از فناوری‌های داده‌کاوی را برای کشف الگوهای قانع‌کننده، معنادار و قابل اعتماد ضروری می‌سازد. کلان‌داده‌های دنیای واقعی را می‌توان به صورت ساختاریافته، بدون ساختار، نیمه‌ساختاریافته و نامتعادل طبقه‌بندی کرد. اهمیت کلاس‌های اکثریت و اقلیت در یک مجموعه داده نامتعادل به مسئله تحقیق بستگی دارد. در این فصل یک روش پیشنهادی برپایه یادگیری ماشین با تلفیق الگوریتم تکاملی جستجوی کلاغ و شبکه عصبی کانولوشن برای ارزیابی و پیش‌بنی ریسک اعتباری ارایه می‌کند.

شکل ۱ مراحل مختلف در چارچوب پیشنهادی را نشان داده است. روش پیشنهادی با پرداختن به چالش‌های ذکر شده قبلی، پویایی‌ها و ویژگی‌های فعلی شرکت‌های کوچک و خرد را ارزیابی می‌کند. با شروع از اطلاعات مالی، اطلاعات خرد رفتاری شرکت‌ها، اطلاعات اعتبار عمومی و دسترسی شخص ثالث به اطلاعات، این روش مجموعه داده معتبر را گردآوری می‌کنیم که هر کدام شامل ۲۰ عنصر اطلاعاتی متمایز هستند. از نظر روش‌شناسی، این تحقیق چارچوبی یکپارچه ایجاد می‌کند که اصلاح داده‌های نامتوازن، تأکید بر ویژگی‌های استراتژیک و روش‌های پیشرفته یادگیری ماشین را با هم ترکیب می‌کند. با کاهش ابعاد داده‌ها و ایجاد یک مدل یادگیری ماشین متناسب، ما نه تنها دقت پیش‌بینی را افزایش می‌دهیم، بلکه اعتبار شرکت‌های کوچک و خرد را از طریق امتیاز اعتباری کمّی می‌کنیم (شکل ۱ را ببینید).

تنوع مجموعه داده‌ها: مجموعه داده‌های ما به دلیل منحصر به فرد بودن، شامل اطلاعات مالی، اطلاعات خرد رفتاری شرکت‌ها، اطلاعات اعتبار عمومی و دسترسی شخص ثالث به اطلاعات، متمایز است.



شکل ۱: فلوچارت روش پیشنهادی

این مجموعه داده جامع، ارزیابی دقیق‌تری از ریسک‌های اعتباری مرتبط با شرکت‌های کوچک و خرد را تسهیل می‌کند. در روش پیشنهادی، تمرکز بر ابعاد اطلاعاتی متنوع شرکت‌های کوچک و خرد و ارزیابی ریسک اعتباری آنها از دیدگاه‌های مختلف ضروری است.

روش‌شناسی نوآورانه: در این گام با استفاده از الگوریتم‌های پردازش نمونه نامتوازن، به چالش‌های نمونه‌های کلاس اقلیت کم‌نمایندگی‌شده می‌پردازیم. علاوه بر استفاده از الگوریتم تکاملی جستجوی کلاغ، از طریق طبقه‌بندی‌کننده‌های پیشرفته یادگیری ماشین، عوامل تعیین‌کننده حیاتی مؤثر بر کاهش اعتبار شرکت‌های کوچک و خرد را مشخص می‌کنیم. با به‌کارگیری رویکرد خود، یک مدل کارت امتیازدهی شبکه عصبی کانولوشن طراحی می‌کنیم که ریشه در این عوامل مهم توصیف‌کننده دارد. این استراتژی مدل، چشم‌انداز جدیدی را به‌ویژه در پرداختن به چالش‌های تحلیل داده‌های با ابعاد بالا، ارائه می‌دهد. استفاده از چنین رویکرد مدل ترکیبی، کاهش ابعاد را تسهیل کرده و پیچیدگی تخمین پارامتر را ساده می‌کند.

ارتباط عملی: یافته‌های در دست فراتر از مشارکت‌های نظری، کاربردهای قابل توجهی در دنیای واقعی ارائه می‌دهند. آنها به عنوان منبعی ارزشمند برای نهادهایی که در تقویت رشد قوی شرکت‌های کوچک و خرد، افزایش مکانیسم‌های هشدار ریسک اعتباری و بهینه‌سازی کارایی تأمین مالی نقش دارند، عمل می‌کنند. علاوه بر این، تحقیقات بعدی می‌تواند بر نوآوری مدل بر اساس شناسایی ویژگی‌های مهم تمرکز کند. نوآوری‌های بالقوه روش پیشنهادی به شرح زیر است:

الگوریتم بهینه‌سازی کلاغ[1]: در این الگوریتم، گله کلاغ ها پخش می شوند و در سراسر فضای تصمیم گیری برای نقاط مخفیگاه کامل (بهینه جهانی) جستجو می کنند. از آنجایی که هر الگوریتم بهینه سازی کارآمد باید با ابعاد دلخواه سازگار باشد و هر بعد دلخواه باید یک متغیر تصمیم را نشان دهد، یک محیط d بعدی برای فضای جستجو در نظر گرفته شده است. در ابتدا، فرض بر این است که افراد کلاغ N (اندازه گله) به طور تصادفی موقعیتی را در فضای d بعدی اشغال می کنند. موقعیت تک کلاغ در تکرار t در فضای جستجو با x(i,t) نشان داده می شود، که در واقع آرایه ای از متغیرهای تصمیم گیری امکان پذیر است. به‌علاوه، هر کلاغ می‌تواند بهترین مکان مخفی‌گاهی را که با آن مواجه شده است، حفظ کند. در تکرار t، موقعیت مخفیگاه فرد کلاغ چهارم با m(i,t) نشان داده می‌شود که بهترین موقعیتی است که تا به حال فرد کلاغ دوم مشاهده کرده است. یک الگوریتم فراابتکاری کارآمد باید تعادل خوبی بین تنوع و تشدید ایجاد کند. در الگوریتم تکاملی کلاغ، تشدید و تنوع عمدتاً توسط دو پارامتر کنترل می شود: طول پرواز (fl) و احتمال آگاهی. با کاهش احتمال آگاهی، شانس تشخیص نقاط مخفیگاه توسط اعضای گله کلاغ افزایش می یابد. در نتیجه، الگوریتم بهینه سازی کلاغ تمایل دارد جستجو را در مجاورت نقاط مخفیگاه متمرکز کند. بنابراین، می توان فرض کرد که مقادیر کوچکتر AP جنبه تشدید الگوریتم را تقویت می کند. از سوی دیگر، با افزایش

[1] Crow Search Algorithm

احتمال آگاهی، گله کلاغ‌ها به احتمال زیاد فضای تصمیم‌گیری را به صورت تصادفی جستجو می‌کنند، در واقع چنین اقدامی شانس کشف نقاط مخفیگاه واقعی توسط غارتگران را کاهش می‌دهد. در نتیجه، مقادیر بزرگتر AP جنبه تنوع بخشی الگوریتم را تقویت می کند.

شبکه عصبی کانولوشن: شبکه عصبی کانولوشن نوع خاصی از شبکه عصبی مصنوعی است. نسخه مدرن یک شبکه عصبی کانولوشن برای اولین بار توسط باتو و لکون[1] [۹] در سال ۱۹۹۸ معرفی شد. نویسندگان ادعا می کنند که شبکه‌های عصبی کانولوشن از نحوه درک بصری پستانداران از جهان با استفاده از معماری نورونها الهام گرفته شده اند. کار آنها از تحقیقات هوبل و ویزل [۱۰] الهام گرفته شده است که در آن، آنها معماری ادراک بصری پستانداران را با جزئیات توصیف می کنند.

معماری شبکه عصبی کانولوشن: شبکه‌های عصبی کانولوشنال سه ایده معماری را با هم ترکیب می کنند. دو ایده اول بخش‌های پذیرای محلی و وزن‌های مشترک عدم تغییر را هنگام اعمال درجه‌ای، مقیاس و اعوجاج را تضمین می‌کنند. و ایده آخر نمونه گیری فرعی، به تنظیمات مکانی یا زمانی اشاره دارد[۱۱]. یک نورون کانولوشنال فقط با چند نورون از لایه قبلی متصل است. به دلیل این اتصال محدود، یک لایه کانولوشن روی یک پچ خاص از ورودی تمرکز می کند. ظرفیت یک شبکه عصبی کانولوشن را می توان با عمق و تنفس متفاوت کنترل کرد. شبکه‌های عصبی کانولوشن به خوبی روی تصاویر کار می کنند، زیرا آنها پیش فرض های خوبی در مورد ویژگی های تصاویر مانند وابستگی های پیکسل محلی و ثابت بودن آمار ایجاد می کنند[۹].

ورودی‌ها: شبکه‌های عصبی کانولوشن به طور کلی برای تصویر استفاده می شوند، هر تصویری را می توان به عنوان ماتریسی از مقادیر پیکسل بیان کرد.

لایه‌ها: برخلاف شبکه‌های پیشخور، یک شبکه عصبی کانولوشن از لایه‌های بسیار متفاوت‌تری تشکیل شده است:

لایه کانولوشن: این لایه خروجی نورون را محاسبه می کند.

لایه ادغام: یک نمونه برداری کوچک از ابعاد فضایی را انجام می دهد. با این کار حجم کمتری از داده های ورودی تولید می شود.

کاملاً متصل: مانند شبکه عصبی پیشخور، هر نورون در یک لایه کاملا متصل به تمام نورون های لایه قبلی متصل می شود.

لایه کانولوشنال: یک لایه کانولوشن عمدتاً برای تصاویر و فیلم ها استفاده می شود. هر دوی آنها ابعاد بالایی دارند و بنابراین اتصال نورون ها به همه نورون های قبلی مانند یک لایه کاملاً متصل غیرعملی است.

سه فراپارامتر وجود دارد که اندازه حجم خروجی را کنترل می کنند: عمق، گام و لایه صفر. عمق مربوط به تعداد فیلترها است که هر کدام چیزی متفاوت از ورودی یاد می گیرند. به عنوان مثال، برخی ممکن است به دنبال لبه‌های متفاوت یا حباب های رنگی باشند. گام ۱ این است که فیلتر چند پیکسل اسلاید است. هنگامی که اسلاید با ۱ مطابقت دارد، فیلتر هر بار یک پیکسل جابجا می شود. اگر گام ۲ باشد، پرش ۲ پیکسلی وجود خواهد داشت. لایه صفر مربوط به اضافه کردن صفر به مرز ورودی است. این اجازه می دهد تا اندازه فضایی حجم خروجی را کنترل کند[۱۲].

لایه کاملاً متصل: مانند شبکه عصبی پیشخور، یک لایه کاملاً متصل دارای اتصالات کامل به تمام فعال‌سازی‌های لایه قبلی است[۱۲].

لایه ادغام: یک روش معمول استفاده از یک لایه ادغام بین لایه‌های کانولوشن است. هدف از یک لایه ادغام کاهش اندازه فضایی است، این نه تنها پارامترها را کاهش می دهد، بلکه محاسبات مورد نیاز را نیز کاهش می دهد،

کاهش بیش‌برازش: معماری های عمیق معمولا دارای میلیون ها پارامتر هستند، هر لایه با اتصالات وزنی به هم متصل می شود.

تنظیم لاسو/ریج[2]: منظم‌سازی لاسو و ریج یک عبارت جریمه اضافه می‌کند که از تطبیق کامل ضریب با داده‌ها جلوگیری می‌کند و در نتیجه از تطبیق بیش از حد جلوگیری می‌کند.

[1] LeCun and Bottou

[2] L1/L2

دراپ‌اوت[1]: تکنیکی است که از برازش بیش از حد جلوگیری می‌کند.

۴- نتایج و ارزیابی روش پیشنهادی

موضوع اصلی این تحقیق پیش‌بینی و ارزیابی ریسک اعتباری مشتریان بانکی است، بهره‌گیری از داده‌های واقعی و متناسب با حوزه بانکداری و اعتبارسنجی ضروری به نظر می‌رسد. در همین راستا، مجموعه داده‌ای که از پایگاه داده کگل انتخاب شده، اطلاعات ارزشمندی را در زمینه ویژگی‌ها و برچسب‌های مربوط به وضعیت اعتباری مشتریان فراهم می‌آورد که می‌تواند به طور مؤثری در توسعه مدل‌های پیشنهادی تحقیق مورد استفاده قرار گیرد. در این فصل، به پیاده‌سازی عملی روش پیشنهادی برای پیش‌بینی و ارزیابی ریسک اعتباری مشتریان بانکی با استفاده از ترکیب الگوریتم تکاملی کلاغ و شبکه عصبی کانولوشن پرداخته می‌شود. هدف اصلی این فصل، ارزیابی کارایی و دقت مدل پیشنهادی در طبقه‌بندی مشتریان بانکی بر اساس سطح ریسک اعتباری آن‌ها است. ارزیابی ریسک اعتباری به عنوان یکی از چالش‌های اساسی در صنعت بانکداری، نیازمند استفاده از مجموعه داده‌های جامع و نماینده است که بتواند ویژگی‌های مختلف مشتریان و الگوهای پیچیده موجود در رفتار پرداخت آن‌ها را منعکس کند. در این راستا، انتخاب مجموعه داده مناسب و تحلیل دقیق آن، نقش بسزایی در موفقیت مدل‌های یادگیری ماشین دارد. ارزیابی ریسک اعتباری به‌عنوان یکی از چالش‌برانگیزترین حوزه‌های تصمیم‌گیری مالی، نقش تعیین‌کننده‌ای در پایداری نظام بانکی و کاهش مطالبات غیرجاری دارد. در این راستا، به‌کارگیری مدل‌های ترکیبی هوشمند نظیر تلفیق الگوریتم‌های تکاملی (مانند الگوریتم کلاغ) با شبکه‌های عصبی عمیق همانند شبکه عصبی کانولوشن قابلیت یادگیری الگوهای پیچیده داده‌های اعتباری را افزایش داده و پیش‌بینی‌های دقیق‌تری ارائه می‌دهد. فصل چهارم این پایان‌نامه، تمرکز خود را بر ارزیابی عملکرد این مدل ترکیبی روی یک مجموعه داده واقعی بانکی متمرکز نموده است. صحت‌سنجی نتایج در محیط عملیاتی، مستلزم به‌کارگیری داده‌های معتبر، جامع و نمایانگر چالش‌های واقعی صنعت بانک‌داری است. در این فصل، ابتدا مجموعه داده انتخابی به‌طور مبسوط معرفی می‌شود تا خواننده درک روشنی از ساختار، مؤلفه‌ها و محدودیت‌های آن کسب نماید.

مجموعه داده: این مجموعه داده ها[2] پیش‌بینی وضعیت وام، داده‌های متقاضیانی را داریم که قبلاً بر اساس ملک که وام ملکی است، برای وام درخواست داده‌اند. بانک بر اساس عواملی مانند درآمد متقاضی، مبلغ وام، سابقه اعتباری قبلی، درآمد متقاضی مشترک و غیره تصمیم می‌گیرد که به متقاضی وام بدهد یا خیر... هدف ما ساخت یک مدل یادگیری ماشینی برای پیش‌بینی تأیید یا رد وام برای یک متقاضی است.

این مجموعه داده با هدف تحلیل و ارزیابی ریسک اعتباری مشتریان بانکی گردآوری شده است و شامل اطلاعات مربوط به وضعیت تسهیلات دریافتی، تسویه یا نکول آن‌ها، ویژگی‌های جمعیت‌شناختی، مالی و سوابق اعتباری مشتریان می‌باشد. مجموعه داده مذکور مناسب برای مسائل طبقه‌بندی است و به صورت مسئله باینری (خوب/بد) یا چندکلاسه، وضعیت ریسک اعتباری را برچسب‌گذاری نموده است. مشخصات کلیدی مجموعه داده:

- **تعداد رکوردها:** ۱۰۰۰۰ نمونه
- **تعداد ویژگی‌ها:** ۲۰ متغیر توصیفی و یک متغیر هدف
- **ویژگی‌های مهم:** سن مشتری، وضعیت تأهل، درآمد ماهانه[3]، مبلغ وام دریافتی[4]، مدت باز پرداخت[5]، سابقه نکول[6]، نسبت بدهی به درآمد[7]، شغل مشتری[8]، وضعیت مسکن[9]، تعداد اقساط پرداخت‌شده[10] و کشور یا منطقه[11].
- **متغیر هدف یا برچسب[12]**: متغیر هدف معمولاً نمایشگر وضعیت اعتباری مشتری است که به صورت باینری)مثلاً خوب/یا بد، پیش‌فرض/غیر پیش‌فر و یا چندکلاسه، ریسک کم/متوسط/زیاد کدگذاری می‌شود.

[1] Dropout
[2] https://www.kaggle.com/datasets/zaurbegiev/my-dataset
[3] Monthly Income
[4] Loan Amount
[5] Repayment Period
[6] Number of Defaults
[7] Debt-to-Income Ratio
[8] Occupation
[9] House Ownership
[10] Installments Paid
[11] Region/Country
[12] Target/Label

داده‌های فوق به صورت جدول "مقادیر جدا شده با کاما[1]" در دسترس است و نسبت به میزان کاملی و عدم وجود داده‌های مفقود[2] بررسی می‌شود.

۵- ارزیابی و شبیه سازی

قابلیت‌اطمینان و اعتبار داده‌های ترافیکی نقش مهمی در سیستم‌های حمل و نقل هوشمند دارند. اکثر طرح‌های شناسایی داده سطح شبکه را هدف قرار می‌دهند، و به ندرت در مورد تاثیر شناسایی داده‌های خودرو در تقاطع‌ها بر کنترل هوشمند چراغ‌های راهنمایی رانندگی بحث می‌کنند. علاوه بر این، اکثر رویکردهای شناسایی داده در یک مرکز ابری متمرکز پردازش می‌شود، که برای تقاطع‌های پیچیده و متغیر قابل‌اجرا نیست. در این خصوص، در این رساله یک طرح شناسایی داده مبتنی بر محاسبات لبه را برای چراغ راهنمایی تقاطع‌ها پیشنهاد می‌کنیم.

در طرح پیشنهادی، چراغ‌های راهنما به عنوان گره‌های لبه برای تشخیص داده‌های خودرو عمل می‌کنند. برای این منظور، سناریوی تقاطع واحد را در نظر می‌گیریم. با کمک ارتباطات V2E، داده‌های خودرو را از ایستگاه پایه جمع‌آوری می‌کنیم، و از فیلتر کویشنت برای تایید قابلیت‌اطمینان و صحت داده‌ها استفاده می‌کنیم. در ادامه، سناریوی تقاطع‌های متعدد را در نظر می‌گیریم. داده‌های خودرو دو تقاطع مجاور را با هم ادغام می‌کنیم و سپس، قابلیت اطمینان داده‌ها را توسط فیلتر کویشنت تایید می‌کنیم. علاوه بر این، این طرح از توابع هش mmh3 موجود در فیلتر کویشنت QF برای کاهش فضای اشغال‌شده منابع محاسباتی گره‌های لبه و نرخ خطای بیتی استفاده می‌کند. نتایج نشان می‌دهند که استفاده از رویکرد شناسایی داده در تشخیص سریع داده‌ها حتی با وجود تعداد زیاد خودرو و داده‌های پیچیده آنها موثر است. طرح پیشنهادی قابلیت‌اطمینان و اثربخشی خودورها را با تاخیر کمتری اثبات نموده است.

اگرچه شبیه‌سازی‌های عددی در مقیاس بزرگ و آزمایش‌های مقیاس کوچک نتایج امیدوارکننده‌ای را نشان داده‌اند، یک درک نظری جامع برای روان کردن جریان ترافیک از طریق داده‌های جمع‌آوری شده وجود ندارد. در اینجا، از دیدگاه تئوری کنترل، ما نتایج تحلیلی را در مورد کنترل پذیری، پایداری، و قابلیت دسترسی یک سیستم ترافیک مختلط متشکل از داده‌ها مربوط به آن در یک جاده با تقاطع واحد و یا چندتایی ایجاد می‌کنیم. با توجه به محدودیت توانایی های ارتباطی در عمل، منابع فقط می تواند اطلاعات جزئی از سیستم ترافیک جهانی را برای بازخورد خود دریافت کند. بنابراین، در نظر گرفتن اطلاعات محلی موجود از وسایل نقلیه همسایه مهم است. این منجر به مفهوم طراحی کنترلر ساخت یافته می شود. در جریان ترافیک مختلط، جوخه های غالب چند تقاطع تنها انتخاب برای تشکیل تعاونی نیست.

استراتژی‌های کنترلی در سازندهای مختلف دوباره طراحی شده و شکل‌گیری بهینه چندین منبع را با استفاده از دیدگاه بهینه‌سازی عملکرد مجموعه بررسی شده است. دو سازنده بهینه غالب، یعنی توزیع یکنواخت و تشکیل جوخه، از آزمایش‌های عددی گسترده پدیدار می‌شوند. با استفاده از ترکیب فیلتر کویشنت و الگوریتم تکاملی ازدحام ذرات که در روش پیشنهادی مورد استفاده قرار گرفته است در نهایت بهبود ۲۰ میلی ثانیه در نتایج به ثبت رسیده است. به جهت بهبود و کارهای آینده پیاده سازی و ارایه روش پیشنهادی در دنیای واقعی و استفاده از فیلتر و یا الگوریتم‌های تکاملی دیگر به جهت بهبود و افزایش کارایی را می‌توان پیشنهاد داد.

برخی ویژگی‌ها نیازمند پیش‌پردازش شامل دسته‌بندی متغیرهای گسسته، نرمال‌سازی متغیرهای عددی و حذف مقادیر پرت[3] می‌باشند که این اقدامات در بخش آتی گزارش خواهند شد. همچنین، توزیع متغیر هدف بررسی شده تا تعادل[4] داده‌ها مورد ارزیابی قرار گیرد. این مجموعه داده به واسطه شفافیت ویژگی‌ها و توصیف وضعیت واقعی مشتریان بانکی، بستری مناسب برای آزمایش الگوریتم‌های پیشرفته یادگیری ماشین، از جمله مدل‌های مبتنی بر شبکه عصبی کانولوشنی و الگوریتم‌های هوشمند تکاملی از جمله الگوریتم کلاغ، فراهم می‌سازد. تنوع بالای ویژگی‌ها، حجم مناسب داده‌ها و مستدل بودن برچسب‌های هدف از نقاط قوت داده حاضر در راستای نیل به اهداف تحقیق محسوب می‌شود.

ارزیابی: در اولین گام از روش پیشنهادی از الگوریتم جستجو کلاغ برای انتخاب سرخوشه استفاده شده است. این الگوریتم برای اجرا باید تعداد تکرار مشخصی را برای اتمام به

[1] Comma-Separated Values
[2] Missing Value
[3] Outlier
[4] Balance

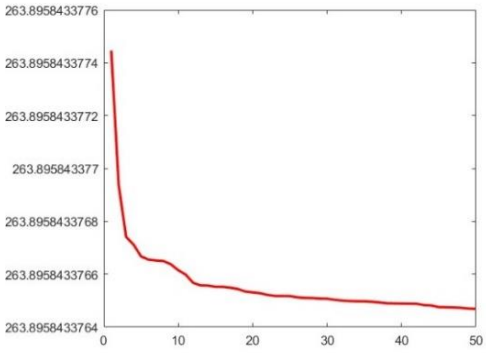
الگوریتم در نظر بگیرد این تعداد باید عددی مناسبی باشد. زیرا این تعداد در صورت زیاد بودن زمان اجرای روش را غیرمنطقی می‌نماید و در صورتی که این تعداد بسیار کم باشد به مقدار بهینه مورد نظر دست نیافته و نتایج نهایی مناسب نخواهند بود. احتمال آگاهی به عنوان یک ایستگاه پایه نامیده می شود، که اختلاف تشدید و تنوع را کاهش می دهد. شبه کد الگوریتم جستجوی کلاغ را برای انتخاب سرخوشه در روش پیشنهادی با پارامترهای مورد استفاده توسط الگوریتم جستجوی کلاغ که در جدول ۱ تعریف شده‌اند به اجرا در خواهند آمد.

جدول ۱: پارامترهای مورد استفاده در الگوریتم جستجوی کلاغ

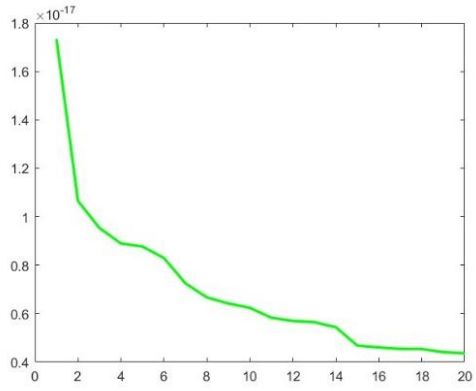
ردیف	پارامتر	شرح آن
۱	Mv	Makespanارزش
۲	TMv	موقتMakespanمقدار
۳	max_iteration	حداکثر تعداد تکرار برای CSA
۴	TAT	جدول تخصیص موقت
۵	AT	جدول توزیع
۶	FL	فاصله پرواز
۷	CT	دوره تکمیل
۸	N	مجموع وظایف
۹	M	ویژگی‌های مجموعه
۱۰	AP	احتمال آگاهی

رویه‌های فرااابتکاری عمدتاً تحت تأثیر فرآیندهای طبیعی هستند و توسط سایر حوزه‌های توسعه انسانی مورد توجه قرار نمی‌گیرند. این یک اتفاق عادی نیست که روش پیشنهادی را تشویق می کند، بلکه یک پدیده اجتماعی-انسانی است. به طور خاص، تولید استعمار در این الگوریتم به عنوان گامی از تکامل سیاسی-اجتماعی بشر ارائه شد و به عنوان منبع الگوریتمی کارآمد برای پالایش این پدیده تاریخی با مدل‌سازی ریاضی مورد استفاده قرار گرفت. از زمان اجرای این الگوریتم برای رفع چندین مشکل در زمینه بهینه‌سازی استفاده شده است. استراتژی های جذب و انقلاب در دو نقش اساسی در داخل روش پیشنهادی مورد استفاده قرار می‌گیرند و نقش اصلی را در الگوریتم ایفا می کنند.

در شکل ۲ نتایج با تعداد تکرار و حد آستانه مختلف در این الگوریتم نشان داده شده است. همانگونه که در شکل ۲ نشان داده شده است مقادیر در تعداد ۲۰ عدد و حد پایین و بالای ۱ و -۱ به همگرایی لازم می‌رسد و نیاز به تکرار بیشتر نیست و یا این تعداد می‌توان به انتخاب مناسب لازم دست یافت بنابراین براساس این شکل‌ها مقدار ۲۰ را به عنوان عدد مناسب تعداد تکرار برمی‌گزینیم. در این مرحله، ویژگی‌های انتخاب‌شده توسط الگوریتم بهینه‌سازی کلاغ به شبکه عصبی کانولوشن داده می‌شود. جدول ۲ مقادیر کمی ارزیابی، شامل دقت[1]، صحت[2] و یادآوری[3] را برای هر Fold در اعتبارسنجی متقابل ۱۰ لایه یادآوری[4] نمایش می‌دهد. همانطور که از میانگین نتایج پیداست، مدل توانسته با دقت بالایی عمل کند



الف) با تعداد ۵۰ تکرار و حد پایین و بالای آزاد



ب) با تعداد ۲۰ تکرار و حد آستانه -۱ و ۱

[1] Precision
[2] Accuracy
[3] Recall
[4] 10-Fold Cross-validation

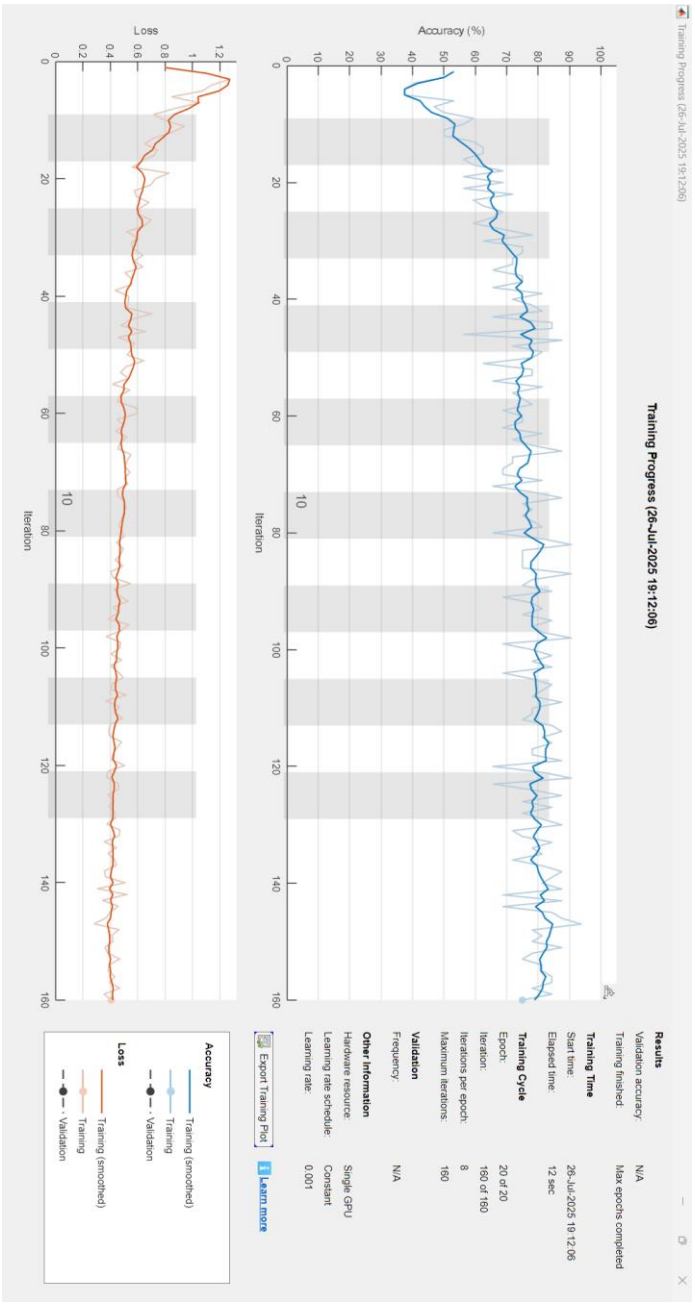
شکل ۲: مقادیر تابع هزینه الگوریتم تکاملی بهینه سازی کلاغ با تعداد تکرار مختلف

شکل ۳ نمودار فرآیند آموزش و اعتبارسنجی مدل شبکه عصبی کانولوشن را نشان می‌دهد. منحنی دقت آموزش[1] و دقت اعتبارسنجی[2] هر دو روند صعودی داشته و در انتهای فرآیند آموزش به یکدیگر نزدیک می‌شوند. این همگرایی نشان می‌دهد که مدل دچار بیش برازش[3] نشده و توانایی تعمیم‌پذیری خوبی بر روی داده‌های جدید دارد. همچنین، منحنی خطای آموزش و اعتبارسنجی نیز روند نزولی داشته و به پایداری رسیده‌اند که این موضوع نیز بیانگر آموزش موفقیت‌آمیز شبکه است.

جدول ۲: آموزش و صحت سنجی شبکه عصبی کانولوشن در مرحله انتخاب ویژگی برروی مجموعه داده

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:09	53.12%	0.7983	0.0010
7	50	00:00:10	78.12%	0.5236	0.0010
13	100	00:00:11	78.12%	0.4679	0.0010
19	150	00:00:12	78.12%	0.4206	0.0010
20	160	00:00:12	75.00%	0.4060	0.0010

[1] Training Accuracy
[2] Validation Accuracy
[3] Overfitting



شکل ۳: آموزش و صحت سنجی شبکه عصبی کانولوشن در مرحله انتخاب ویژگی برروی مجموعه داده

جدول ۳: آموزش و صحت سنجی شبکه عصبی کانولوشن در مرحله طبقه بندی برروی مجموعه داده

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:02	21.88%	1.2661	0.0010
7	50	00:00:03	78.12%	0.4673	0.0010
13	100	00:00:03	71.88%	0.4533	0.0010
19	150	00:00:04	81.25%	0.3505	0.0010
20	160	00:00:04	81.25%	0.3066	0.0010

این جدول و نمودار، نتایج نهایی مدل ترکیبی پیشنهادی را در طبقه‌بندی مشتریان بانکی نمایش می‌دهند. نتایج حاصل از اعتبارسنجی ۱۰ لایه در جدول ۳ نشان‌دهنده عملکرد قدرتمند مدل در تفکیک مشتریان معتبر از نامعتبر است. شکل ۴ نیز به صورت بصری، پایداری و کارایی بالای مدل را تأیید می‌کند. نزدیکی منحنی‌های آموزش و اعتبارسنجی، بیانگر آن است که مدل پیشنهادی توانسته الگوهای پیچیده موجود در داده‌های اعتباری را به خوبی یاد بگیرد و آن‌ها را به داده‌های دیده‌نشده تعمیم دهد.

در کلیه آزمایش‌ها از تست صحت به روش K-Fold با K=10 استفاده شده است. در این نوع اعتبارسنجی داده ها به K زیرمجموعه افراز می شوند. از این K زیرمجموعه، هر بار یکی برای اعتبارسنجی و K-1 تای دیگر برای آموزش بکار میروند. این روال K بار تکرار می شود و همه داده ها دقیقا یکبار برای آموزش و یکبار برای اعتبارسنجی بکار می روند. در نهایت میانگین نتیجه این K بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده می شود. البته می توان از روش های دیگر برای ترکیب نتایج استفاده کرد. بطور معمول از ۱۰-Fold استفاده می شود.برای ارزیابی از الگوریتم‌های طبقه‌بندی K نزدیکترین همسایه، بهینه‌سازی حداقلی متوالی، الگوریتم مبتنی بر نظریه بیز، J48 و بگینگ استفاده شده است[۱۳] و معیارهای ارزیابی به ترتیب زیر است:

صحت: اثربخشی یک مدل با نسبت پیش‌بینی‌های دقیق تولید شده در تمام انواع پیش‌بینی‌ها ارزیابی می‌شود. فرآیند ارزیابی شامل ارزیابی صحت طبقه‌بندی است یعنی نمونه‌های طبقه‌بندی شده با تعداد کلی رخدادها. اندازه‌گیری صحت به ویژه در مواردی ارزشمند است که توزیع کلاس‌ها در متغیر هدف به طور یکنواخت در سراسر مجموعه داده پخش شود. رابطه صحت در معادله (۱) بیان شده است:[۱۴]

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \quad (۱)$$

یادآوری: یادآوری اندازه‌گیری نرخ مثبت واقعی در زمینه یک سیستم نقص نرم‌افزاری است. در این زمینه خاص، تعداد رخدادهایی را نشان می‌دهد که به عنوان نرم‌افزار معیوب طبقه‌بندی شده‌اند که به‌طور دقیق توسط مدل پیش‌بینی شده‌اند. معادله ۲ نشان دهنده نسبت نمونه‌های نرم‌افزاری مشکل‌ساز است که به طور دقیق توسط مدل شناسایی شده-اند:[۱۴]

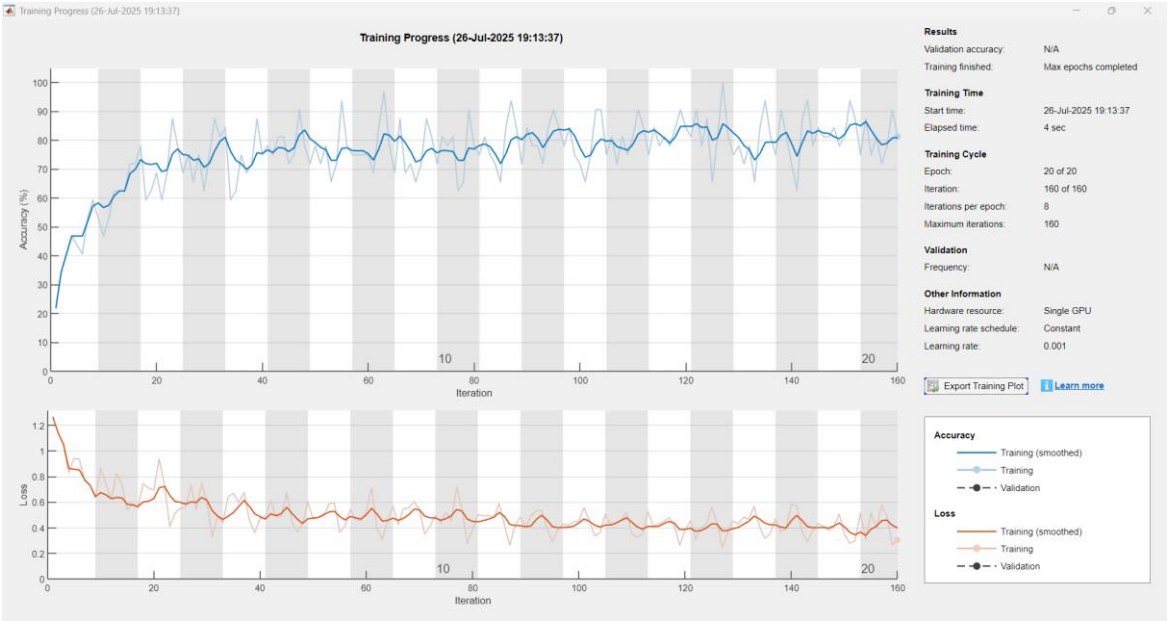
$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (۲)$$

دقت: دقت پیش‌بینی‌های مثبت انجام‌شده توسط مدل را اندازه‌گیری می‌کند و نسبت نمونه‌های پیش‌بینی‌شده به‌عنوان مثبت و مثبت واقعی را نشان می‌دهد و با استفاده از فرمول زیر محاسبه می شود:[۱۵]

$$\text{precision} = \frac{\text{TN}}{\text{TN+FP}} \quad (۳)$$

امتیاز F1: معیاری است که دقت و یادآوری را در یک مقدار واحد ترکیب می‌کند. این به ویژه زمانی مفید است که توزیع کلاس ناهموار (کلاس های نامتعادل) وجود داشته باشد و اغلب در مسائل طبقه بندی باینری استفاده می‌شود. این معیار میانگین هارمونیک دقت و یادآوری است و با استفاده از فرمول زیر محاسبه می شود:[۱۵]

$$\text{F} - \text{Measure} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision+ recall}} \quad (۴)$$



شکل ۴: آموزش و صحت سنجی شبکه عصبی کانولوشن در مرحله طبقه بندی برروی مجموعه داده

Model	Accuracy	Precision	Recall	F1
{'NaiveBayes'}	{[0.7491]}	{[0.5714]}	{[0.4737]}	{[0.5180]}
{'KNN' }	{[0.7603]}	{[0.6429]}	{[0.3553]}	{[0.4576]}
{'J48' }	{[0.6592]}	{[0.3973]}	{[0.3816]}	{[0.3893]}
{'SVM' }	{[0.7790]}	{[0.6308]}	{[0.5395]}	{[0.5816]}
{'Bagging' }	{[0.8015]}	{[0.7255]}	{[0.4868]}	{[0.5827]}

شکل ۵: مقایسه روش های مختلف و روش پیشنهادی

جدول ۴: مقایسه روش های مختلف و روش پیشنهادی

روش	معیار صحت	معیار Precision	معیار Recall	معیار F1
بیزین	۷۴.۹۱	۰.۵۷۱۴	۰.۴۷۳۷	۰.۵۱۸۰
KNN	۷۶.۰۳	۰.۶۴۲۹	۰.۳۵۵۳	۰.۴۵۷۶
J48	۶۵.۹۲	۰.۳۹۷۳	۰.۳۸۱۶	۰.۳۸۹۳
SVM	۷۷.۹	۰.۶۳۰۸	۰.۵۳۹۵	۰.۵۸۱۶
Bagging	۸۰.۱۵	۰.۷۲۵۵	۰.۴۸۶۸	۰.۵۸۲۷
روش پیشنهادی	۸۲.۵۴	۰.۸۲۱۲	۰.۵۶۵	۰.۶۲۴

براساس جدول ۴ روش پیشنهادی به نسبت سایر روشها دارای عملکرد بهتری بوده و به نسبت نزدیکترین روش ۲ درصد بهبود داشته است.

در این نمودار، عملکرد مدل پیشنهادی (الگوریتم تکاملی کلاغ-شبکه عصبی کانولوشن) با سایر روش‌های متداول طبقه‌بندی مانند نظریه بیز و بگینگ بر اساس معیار دقت مقایسه شده است. همانطور که مشاهده می‌شود، روش پیشنهادی به طور محسوسی عملکرد بهتری نسبت به سایر الگوریتم‌ها داشته است. این برتری را می‌توان به دو دلیل عمده نسبت داد:

انتخاب ویژگی هوشمند :الگوریتم کلاغ توانسته است با حذف ویژگی‌های نامرتبط و نویزی، یک زیرمجموعه بهینه از ویژگی‌ها را انتخاب کند که این امر باعث کاهش پیچیدگی و افزایش دقت مدل طبقه‌بندی شده است.

قدرت طبقه‌بندی شبکه عصبی عمیق: شبکه عصبی کانولوشن به دلیل قابلیت بالای خود در استخراج الگوهای پیچیده و غیرخطی از داده‌ها، توانسته است روابط میان ویژگی‌های منتخب را به بهترین شکل یاد گرفته و طبقه‌بندی دقیق‌تری را ارائه دهد. این نتایج نشان می‌دهد که ترکیب یک الگوریتم بهینه‌سازی فراابتکاری با یک مدل یادگیری عمیق، استراتژی مؤثری برای حل مسئله چالش‌برانگیز ارزیابی ریسک اعتباری است.

۶- نتیجه گیری

در این فصل، مروری جامع بر دستاوردهای تحقیق ارائه شده و نتایج کلیدی حاصل از تحلیل داده‌ها و اجرای مدل‌ها خلاصه شده‌اند. همچنین، محدودیت‌ها و چالش‌های موجود معرفی شده و راهکارها و پیشنهاداتی برای بهبود در تحقیقات آینده بیان خواهد شد. این تحقیق گامی موثر در جهت ارتقاء دقت و قابلیت اعتماد در پیش‌بینی ریسک اعتباری مشتریان بانکی برداشته است. با استفاده از مدل ترکیبی الگوریتم تکاملی کلاغ و شبکه عصبی کانولوشن، امکان بهره‌گیری از داده‌های واقعی و پیچیده به نحو بهینه فراهم آمده است. این مدل می‌تواند به عنوان ابزاری کاربردی در بانک‌ها جهت کاهش ریسک مالی و بهبود فرآیندهای تصمیم‌گیری اعتباری بکار گرفته شود. مدل ترکیبی الگوریتم تکاملی کلاغ و شبکه عصبی کانولوشن ، با بهره‌گیری از الگوریتم بهینه‌سازی فراابتکاری برای انتخاب ویژگی و قدرت یادگیری شبکه‌های عمیق کانولوشنی، موفق به خلق مدلی با عملکرد بالاتر و دقت بیشتر شده است. استفاده از اعتبارسنجی K-Flod10 به اطمینان از صحت نتایج کمک کرده است. مقایسه با دیگر الگوریتم‌های شاخص موجود در وکا، اعتبار و کارایی مدل را تایید می‌کند. این مدل قابلیت کاربرد در محیط‌های عملی و واقعی برای پیش‌بینی دقیق‌تر ریسک اعتباری دارد.

خلاصه نتایج تحقیق

- مدل پیشنهادی مبتنی بر تلفیق الگوریتم تکاملی کلاغ و شبکه عصبی کانولوشن، به طور قابل توجهی توانست دقت و کارایی پیش‌بینی ریسک اعتباری را افزایش دهد.
- استفاده از مجموعه داده واقعی با ویژگی‌های متنوع زمینه‌ساز تحلیل‌های جامع و دقیق گردید.
- الگوریتم‌های طبقه‌بندی متداول مانند بیز، بگینگ و بهینه‌سازی متوالی کمینه در مقایسه با مدل پیشنهادی عملکرد پایین‌تری نشان دادند که اهمیت انتخاب ویژگی و مدل‌سازی پیچیده‌تر را برجسته می‌کند.
- ارزیابی معیارهای مختلف عملکرد مدل‌ها نشان داد که مدل پیشنهادی قابلیت تفکیک بهینه بین کلاس‌های ریسک اعتباری را داراست.

محدودیت‌های تحقیق:

- محدود بودن اندازه و تنوع داده‌ها نسبت به برخی کاربردهای واقعی بزرگ‌تر
- پیش‌پردازش داده‌ها و حذف مقادیر پرت که ممکن است اثراتی بر نتایج داشته باشند
- پیچیدگی محاسباتی مدل‌های عمیق و نیاز به سخت‌افزار قوی

با توجه به نتایج امیدوارکننده این تحقیق و ماهیت پویای حوزه یادگیری ماشین، مسیرهای متعددی برای ادامه و گسترش این پژوهش در آینده وجود دارد که در ادامه به برخی از آن‌ها اشاره می‌شود:

- **استفاده از الگوریتم‌های فراابتکاری دیگر برای مدل:** می‌توان عملکرد الگوریتم کلاغ را با سایر الگوریتم‌های بهینه‌سازی جدید و قدرتمند مانند الگوریتم بهینه‌سازی گرگ خاکستری، الگوریتم وال یا الگوریتم بهینه‌سازی ازدحام ذرات برای انتخاب ویژگی مقایسه نمود تا بهترین رویکرد شناسایی شود.
- **کاوش در معماری‌های یادگیری عمیق:** به‌جای شبکه عصبی کانولوشن، می‌توان از سایر معماری‌های پیشرفته یادگیری عمیق مانند شبکه‌های عصبی بازگشتی یا حافظه طولانی کوتاه مدت استفاده کرد، به‌ویژه اگر داده‌ها دارای ویژگی‌های زمانی یا توالی باشند. همچنین، ترکیب مدل‌ها مانند شبکه‌های عصبی بازگشتی-حافظه طولانی کوتاه مدت می‌تواند نتایج را بهبود بخشد.

- **به‌کارگیری مجموعه داده‌های بزرگ‌تر و متنوع‌تر برای ارزیابی گسترده‌تر:** ارزیابی مدل بر روی مجموعه داده‌های بزرگ‌تر، نامتوازن و یا داده‌های اختصاصی یک بانک واقعی می‌تواند به درک عمیق‌تری از قابلیت تعمیم‌پذیری و چالش‌های عملیاتی مدل منجر شود.
- **تمرکز بر تفسیرپذیری مدل:** مدل‌های یادگیری عمیق اغلب به عنوان "جعبه سیاه" شناخته می‌شوند. استفاده از تکنیک‌های هوش مصنوعی می‌تواند به درک اینکه کدام ویژگی‌ها بیشترین تأثیر را در تصمیم‌گیری مدل داشته‌اند کمک کرده و اعتماد به مدل را در محیط‌های عملیاتی افزایش دهد.
- **توسعه یک سیستم بلادرنگ:** با توسعه می‌توان مدل توسعه‌یافته را به یک نرم‌افزار کاربردی یا یک وب‌سرویس تبدیل کرد تا کارشناسان اعتباری در بانک‌ها بتوانند به صورت بلادرنگ از آن برای ارزیابی درخواست‌های وام استفاده کنند.
- **ارزیابی با معیارهای دیگر:** علاوه بر دقت، می‌توان از معیارهای ارزیابی دیگری که برای مسائل بانکی و داده‌های نامتوازن مناسب‌تر هستند، مانند امتیازF1، سطح زیر نمودار و ماتریس درهم‌ریختگی برای تحلیل دقیق‌تر عملکرد مدل بهره برد.

۷- منابع

1. Soni, U. and D.A.G. Gordhan Jethava, Latest Advancements in Credit Risk Assessment with Machine Learning and Deep Learning Techniques. Cybernetics and Information Technologies, 2024. 24(4).

2. Abbas, E.A. and N.A. Hussein, Algorithm Comparison for Data Mining Classification: Assessing Bank Customer Credit Scoring Default Risk. Jurnal Kejuruteraan, 2024. 36(5): p. 1935-1944.

3. Zhang, X. and L. Yu, Consumer credit risk assessment: A review from the state-of-the-art classification algorithms, data traits, and learning methods. Expert Systems with Applications, 2024. 237: p. 121484.

4. Salihu, A. and V. Shehu, Data mining based classifiers for credit risk analysis. Managing Global Transitions, 2020. 18(2): p. 147-167.

5. Obunadike, C., Machine learning for credit risk analysis across the United States. Copyright © 2024 Author(s) retain the copyright of this article., 2024.

6. Gao, M.Y., FINANCIAL CREDIT RISK ASSESSMENT BASED ON MACHINE LEARNING. © By the Author(s) 2024, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0). 2024.

7. Gu, Z., et al., Credit risk assessment of small and micro enterprise based on machine learning. Heliyon, 2024. 10(5).

8. Nasir, F., et al., Data-Driven Decision-Making for Bank Target Marketing Using Supervised Learning Classifiers on Imbalanced Big Data. Computers, Materials & Continua, 2024. 81(1).

9. LeCun, Y., C. Cortes, and C. Burges, MNIST handwritten digit database. AT&T Labs {Online}. MNIST handwritten digit database. AT&T labs [online], 2010.

10.LeCun, Y., et al., Gradient-based learning applied to document recognition. Proceedings of the IEEE, 2002. 86(11): p. 2278-2324.

11.Hubel, D.H. and T.N. Wiesel, Receptive fields and functional architecture of monkey striate cortex. The Journal of physiology, 1968. 195(1): p. 215-243.

12.Karpath., CS231n Convolutional Neural Networks for Visual Recognition,. 2018.

13.Gu, Z., Lv, J., Wu, B., Hu, Z., & Yu, X. (2024). , Credit risk assessment of small and micro enterprise based on machine learning. Heliyon, 10(5). 2024.

14.Kene Tochukwu Anyachebelu, S.H.H., Muhammad Umar Abdullahi, Maimuna Abdullahi Ibrahim, Comparative Analysis of Machine Learning Algorithms for Breast Cancer Prediction Dutse Journal of Pure and Applied Sciences (DUJOPAS), Vol. 9 No. 4b December 2023, 2023.

15.Islam, M.R., M.S. Islam, and S. Majumder, Breast cancer prediction: a fusion of genetic algorithm, chemical reaction optimization, and machine learning techniques. Applied Computational Intelligence and Soft Computing, 2024. 2024(1): p. 7221343.

سحر کریم‌لو سیاح: کارشناس ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیکز از دانشگاه آزاد اسلامی واحد تهران جنوب می‌باشد و نشانه رایانامه ایشان:
sayah.sahar@gmail.com.

روش ارجاع: س. کریم‌لو سیاح. ارائه یک مدل ریسک اعتباری بانکی جهت طبقه‌بندی مشتریان مبتنی بر الگوریتم تکاملی کلاغ و شبکه عصبی کانولوشن. دو فصلنامه محاسبات و سامانه های توزیع شده، سال هشتم، شماره۲، شماره پیاپی۱۶، صفحه ۸۸ تا ۱۰۲، سال ۱۴۰۴.

How to cite: S.karimloo sayah, Presenting a bank credit model based on classification and segmentation based on the crow developmental algorithm and convolutional neural network, Journal of Distributed Computing and Systems (JDCS), Vol 8, Issue 2, Page 88-102, 2025.

Presenting a bank credit model based on classification and segmentation based on the crow developmental algorithm and convolutional neural network

S.karimloo sayah
Islamic Azad University, Jonoob Tehran Branch

Abstract

Creditworthiness is one of the most vital and important tasks in the modern banking industry, which ensures the non-performing and sustainability of credit financial institutions with an accurate prediction of the credit status of loan applicants.The main goal of this research is to improve credit assessment for small companies, which is an innovative measure in this field.In fact, to achieve this goal, comprehensive datasets including a variety of financial information, micro-company information, public credit information, and third-party personal access to information are required so that unbalanced sample processing techniques can be used to achieve fairer samples. In this regard, a crow algorithm is used to solve the feature selection problem and in the second step, the selected features are provided as input to a convolutional neural network for final classification. The proposed model was implemented and evaluated on a real-time validated dataset from the Kaggle platform, which included real banking information. K-Fold cross-validation method was used to accurately measure the model performance. The experimental results show that this hybrid model has achieved significantly higher accuracy compared to traditional classification methods. Finally, this research showed that the proposed model has high potential for use as a decision support system in banks and financial institutions for credit assessment.

Key words: Credit prediction and validation, bank classification, crow algorithm and convolutional neural network