DOR: https://csj.isi.org.ir/20.1001.1.2538161.1405.11.41.4.7

DOI: https://doi.org/10.22034/csj.2026.251331

تابستان ۱۴۰۵، دوره ۱۱، شماره ۲، پیاپی ۴۱

دریافت مقاله: ۱۴۰۴/۰۸/۲۶

پذیرش مقاله: ۱۴۰۴/۰۹/۲۲

نوع مقاله: پژوهشی

کشف تقلب در کارت‌های اعتباری با استخراج ویژگی از شبکه عصبی و طبقه‌بندی تقویت گرادیان سبک وزن

مائده نصیری

دانشجو، گروه ریاضیات و کاربردها، دانشکده علوم پایه، دانشگاه شاهد، تهران.

پست الکترونیکی: maedenasiri2000@gmail.com

ابوالفضل تاری مرزآباد*

دانشیار، گروه ریاضیات و کاربردها، دانشکده علوم پایه، دانشگاه شاهد، تهران.

پست الکترونیکی: tari@shahed.ac.ir

چکیده

با افزایش اســتفاده از کارت‌های اعتباری، تشــخیص خودکار تراکنش‌های متقلبانه اهمیت بیشــتری یافته است. در ایــن پژوهش، دو مدل ترکیبی مبتنی بر شــبکه عصبی پیچشی و پرسپترون چندلایه برای استخراج ویژگی به‌کار گرفته شــد و ویژگی‌های استخراج‌شده به الگوریتم تقویت گرادیان سبک‌وزن منتقل گردید. برای انتخاب بهترین لایه، پنج معیار کمّی شــامل شــاخص‌های کالینسکی- هاراباز، دیویس- بولدیــن، اطلاعات متقابل، میزان افزونگی و ابعاد لایه برای همه لایه‌ها محاسبه و لایه بهینه براساس میانگین نرمال‌شــده معیارها تعیین شد. نتایج نشــان داد مدل‌های ترکیبی عملکرد بهتری نســبت به مدل منفرد دارند و مدل مبتنی بر شبکه پیچشی بیشترین کارایی را ارائه کرده است. این مدل موجب افزایش امتیاز F1، کاهش قابل‌توجه خطای مثبت کاذب و افزایش اندک خطای منفی کاذب شد؛ اما این افزایش در برابر کاهش محسوس خطای مثبت کاذب قابل پذیرش بوده و تعادل میان دقت و بازیابی را حفظ می‌کند. آزمون ویلکاکسون تفاوت عملکرد را از نظر آماری معنادار نشــان نداد، با این حال بهبود معیارهای عملکردی اهمیت اســتفاده از مدل ترکیبی مبتنی بر شبکه هم‌آمیختی را در سناریوهای واقعی تشخیص تقلب برجسته می‌سازد. تحلیل SHAP نیز نشــان داد ویژگی دوم استخراج‌شــده از شبکه هم‌آمیختی بیشــترین نقش را داشــته و تحت تأثیر متغیر «مقدار تراکنش» شکل گرفته است.

کلید واژه‌ها: کشــف تقلب، مدل‌های ترکیبی، استخراج ویژگی، شبکه عصبی، تقویت گرادیان سبک وزن

۱- مقدمه

مهاجــرت کســب و کارهــا بــه اینترنت و گســترش تراکنش‌های الکترونیکی در اقتصاد بدون پول نقد که به‌طور مداوم در حال رشد است[۱]، باعث افزایش چشم‌گیر حجم تراکنش‌ها شده است. این موضوع همچنین پیچیدگی و تنوع روش‌های تقلب را افزایش داده و مقابله با آن را دشوارتر کرده است. شرکت‌های صادرکننده کارت اعتباری معمولاً

* نویسندهٔ مسئول

آمار خسارت‌ها را منتشر نمی‌کنند. بنابراین، دستیابی به برآورد دقیق خسارت‌ها دشوار است. با این حال، برخی داده‌های عمومی درباره زیان‌های مالی ناشی از تقلب موجود است.

استفاده از کارت‌های اعتباری بدون امنیت کافی، میلیاردها دلار خسارت مالی به‌همراه دارد[۲]. در گذشته، سامانه‌های تشخیص تقلب با تکیه بر معیارهای از پیش تعریف‌شده، تراکنش‌های مشکوک را شناسایی می‌کردند. با این حال، محدودیت این سامانه‌ها در انطباق با الگوهای نوظهور و نرخ بالای هشدارهای کاذب، عملکرد آنها را تضعیف کرد. در نتیجه، رویکردهای مبتنی بر یادگیری ماشین و یادگیری عمیق با هدف ارتقای دقت و تطبیق‌پذیری[۱] به‌تدریج جایگزین روش‌های سنتی شدند[۳]. این الگوریتم‌ها قادرند با بهره‌گیری از داده‌های حجیم و تحلیل الگوهای رفتاری پیچیده، ناهنجاری‌هایی را که ممکن است نشان‌دهنده فعالیت متقلبانه باشند، شناسایی کنند[٤].

یکی از رویکردهای مطرح و مؤثر در این زمینه، بهره‌گیری از الگوریتم‌های یادگیری ماشین برای طبقه‌بندی تراکنش‌های مشکوک است. در سال‌های اخیر، تحقیقات زیادی در مورد استخراج ویژگی از شبکه‌های عصبی و ترکیب آن با الگوریتم‌های یادگیری ماشین برای افزایش دقت در کشف تقلب انجام گرفته است. با این حال، بررسی مطالعات موجود نشان می‌دهد که برخی چالش‌های مهم همچنان بدون پاسخ باقی مانده‌اند، از جمله:

- عدم توجه به انتخاب لایه بهینه برای استخراج ویژگی از شبکه‌های عصبی که می‌تواند نقش تعیین‌کننده‌ای در عملکرد مدل داشته باشد.
- فقدان تحلیل آماری دقیق برای ارزیابی معناداری تفاوت عملکرد بین مدل‌ها.
- کم‌توجهی به تفسیرپذیری مدل‌ها و تحلیل علل تصمیم‌گیری آنها، که می‌تواند بینش مهمی درباره رفتار مدل ارائه دهد.

پژوهش حاضر با هدف رفع این شکاف‌ها، چارچوبی ترکیبی برای تشخیص تقلب ارائه می‌دهد. در این چارچوب، از شبکه عصبی هم‌آمیختی[۲] و شبکه عصبی پرسپترون چند لایه[۳] برای استخراج ویژگی‌های عمیق استفاده شده و طبقه‌بندی با الگوریتم تقویت گرادیان سبک وزن[٤] انجام می‌شود. همچنین برای ارزیابی دقیق عملکرد مدل، از تحلیل‌های آماری و ابزارهای تفسیرپذیری بهره گرفته‌ایم.

ساختار این مقاله به‌شرح زیر است:

در بخش دوم، مرور ادبیات پژوهش ارائه می‌شود. بخش سوم به معرفی مجموعه‌داده و مدل‌های پایه اختصاص دارد. در این بخش، فرایند پیش‌پردازش، معیارهای ارزیابی عملکرد، ساختار شبکه‌های عصبی و الگوریتم تقویت گرادیان سبک‌وزن تشریح می‌شوند. روش پیشنهادی در بخش چهارم توصیف می‌شود. نتایج تجربی در بخش پنجم گزارش و تحلیل شده و تفسیر نتایج مدل در بخش ششم مورد بررسی قرار می‌گیرد. در نهایت، بخش هفتم به نتیجه‌گیری و ارائه پیشنهادهایی برای تحقیقات آینده اختصاص دارد.

۲- پیشینه تحقیق

در یکی از مطالعات اخیر روشی ترکیبی با ترکیب رمزگذار خودکار[۵] و جنگل تصادفی احتمالاتی[٦] برای تشخیص تقلب در

تراکنش‌های کارت اعتباری ارائه شده است. در این روش، ابتدا با استفاده از رمزگذار خودکار ابعاد داده‌ها کاهش یافته و ویژگی‌های مهم استخراج می‌شوند، سپس با استفاده از جنگل تصادفی با طبقه‌بندی احتمالات تراکنش‌ها به‌صورت عدم تقلب یا تقلب طبقه‌بندی می‌شود[۵].

در [٦]، با به‌کارگیری ترکیب شبکه عصبی هم‌آمیختی به‌منظور استخراج مؤثر ویژگی‌ها و ماشین بردار پشتیبان[۷] به‌عنوان طبقه‌بند نهایی، چارچوبی ارائه شده است که در

1-Adaptability

2-Convolutional neural network
3-Multilayer perceptron
4-Lightweight Gradient Boosting Machine (LightGBM)
5-Autoencoder
6-Probabilistic Random Forest
7-Support vector machines

مقایسه با رویکردهای متداول، دقت بالاتری در تشخیص تقلب در تراکنش‌های کارت اعتباری ارائه می‌دهد.

در پژوهشی دیگر، مدلی ترکیبی مبتنی بر شبکه عصبی هم‌آمیختی، حافظه طولانی کوتاه مدت[8] و الگوریتم k–نزدیک‌ترین همسایه[9] برای تشخیص تقلب در تراکنش‌های بانکی ارائه شده است. مدل پیشنهادی با صحت ۹۹/۵ درصد، میزان بازیابی ۹۲/۱ درصد و مساحت زیر منحنی مشخصه عملکرد گیرنده[10] برابر با ۹۷/۵ درصد، عملکرد بهتری نسبت به الگوریتم‌های سنتی مانند رگرسیون لجستیک[11]، جنگل تصادفی و درخت تقویت‌شده تدریجی[12] دارد. ترکیب لایه‌های شبکه عصبی هم‌آمیختی برای درک روابط پیچیده میان ویژگی‌ها و لایه‌های حافظه بلندمدت–کوتاه‌مدت برای تحلیل الگوهای زمانی، موجب شناسایی دقیق‌تر تراکنش‌های متقلبانه شده است. این مدل، علی‌رغم پیچیدگی و چالش‌هایی در زمینه تفسیرپذیری، راهکاری نویدبخش برای استفاده در سامانه‌های مالی ارائه می‌دهد[۷].

در [۸] با تلفیق شبکه عصبی هم‌آمیختی به‌عنوان استخراج‌کننده ویژگی و الگوریتم جنگل تصادفی به‌عنوان طبقه‌بند، یک مدل ترکیبی برای تشخیص تقلب در تراکنش‌های کارت اعتباری ارائه شده است. نتایج تجربی نشان می‌دهند که این مدل با دستیابی به صحت ۹۸/ ۹۹ درصد، عملکردی فراتر از سایر الگوریتم‌های کلاسیک مانند ماشین بردار پشتیبان با حداکثر صحت ۹۹/۹٤ درصد داشته است.

در [۹]، روشی خلاقانه برای تشخیص تقلب در کارت‌های اعتباری ارائه شده است که از ترکیب یادگیری عمیق و بینایی ماشین بهره می‌برد. در این رویکرد، داده‌های متنی تراکنش‌ها با استفاده از تکنیکی نو به نام تبدیل متن به تصویر[13] به تصاویر کوچکی تبدیل شده و سپس به یک معماری شبکه عصبی هم‌آمیختی تغذیه می‌شوند. برای مقابله با عدم تعادل شدید کلاس‌ها، از روش وزن‌دهی معکوس فرکانس[14] استفاده شده است. نتایج تجربی نشان می‌دهند که ویژگی‌های عمیق استخراج‌شده از شبکه عصبی هم‌آمیختی، زمانی که به الگوریتم‌های کلاسیک یادگیری ماشین مانند نزدیک‌ترین همسایه خشن[15] داده می‌شوند، منجر به دقتی معادل ۹۹/۸۷ درصد در تشخیص تراکنش‌های متقلبانه شده‌اند. این روش با معرفی بعدی جدید در تبدیل داده‌های متنی به تصاویر، مسیرهای جدیدی برای به‌کارگیری تکنیک‌های بینایی ماشین در حوزه تشخیص تقلب پیشنهاد می‌دهد.

در [۱۰]، مدلی ترکیبی دیگری برای تشخیص تقلب در تراکنش‌های کارت اعتباری معرفی شده است که از ادغام پرسپترون چندلایه به‌عنوان استخراج‌کننده ویژگی و طبقه‌بند جنگل تصادفی بهره می‌برد. هدف این مدل، بهره‌گیری هم‌زمان از قدرت یادگیری الگوهای پیچیده توسط پرسپترون چندلایه و قابلیت تصمیم‌گیری قوی جنگل تصادفی است. پس از پیش‌پردازش داده‌ها و متعادل‌سازی آنها با استفاده از تکنیک نمونه‌برداری افزایشی مصنوعی برای کلاس اقلیت، عملکرد مدل پیشنهادی با مدل‌های مستقل پرسپترون چندلایه و جنگل تصادفی مقایسه شده است. نتایج نشان می‌دهد که مدل ترکیبی عملکرد بهتری نسبت به مدل‌های مستقل دارد و با صحت ۹۹/۹٤ درصد، دقت ۸۷/۰۹ درصد، بازیابی ۸۲/٦۵ درصد و امتیاز F1 معادل ۸٤/۸۱ درصد، توانسته است بهبود قابل‌توجهی در تشخیص تراکنش‌های متقلبانه ایجاد کند.

در [۱۱]، برای تشخیص تقلب، ویژگی‌های استخراج‌شده از شبکه عصبی هم‌آمیختی به‌عنوان ورودی به مدل‌های یادگیری ماشین داده شدند. مدل‌های یادگیری ماشین شامل بیز ساده[16]، k–نزدیک‌ترین همسایه، درخت تصمیم[17] و ماشین بردار پشتیبان به‌صورت تجمعی با ویژگی‌های

8-Long Short-Term Memory
9-K-Nearest Neighbor
10-Area Under the Curve
11-Logistic regression
12-Gradient boosting
13-Text2IMG

14-Inverse Frequency Weighting
15-Coarse-KNN
16-Naïve Bayes
17-Decision Tree

جدول ۱. مروری مقایسه‌ای از پژوهش‌های منتخب

دسته‌بندی مدل ترکیبی	مرجع	مجموعه داده	مدل	نوع پیش‌پرازش	لایه استخراج ویژگی
شبکه عصبی هم‌آمیختی + الگوریتم‌های طبقه‌بند	[۶]	داده‌های اروپایی	شبکه عصبی هم‌آمیختی + ماشین بردار پشتیبان	متعادل‌سازی	لایه تخت‌شده
	[۸]	داده‌های اروپایی	شبکه عصبی هم‌آمیختی + جنگل تصادفی	نرمال‌سازی + حذف ویژگی‌های غیرضروری	لایه تخت‌شده
	[۱۱]	داده‌های اروپایی	شبکه عصبی هم‌آمیختی + یادگیری تجمیعی (k_نزدیک‌ترین همسایه، درخت تصمیم، بیزساده، ماشین بردار پشتیبان)	متعادل‌سازی	لایه سوم تمام متصل
	[۹]	داده‌های اروپایی	شبکه عصبی هم‌آمیختی + k-نزدیک‌ترین همسایه	تبدیل داده‌ها تصاویر ۶×۵ پیکسلی + وزن‌دهی معکوس کلاس‌ها	لایه تمام متصل
پرسپترون چندلایه + الگوریتم‌های طبقه‌بند	[۱۰]	داده‌های اروپایی	پرسپترون چندلایه + جنگل تصادفی	پاک‌سازی داده‌ها + نرمال‌سازی + متعادل‌سازی	لایه آخر
چند شبکه عصبی + الگوریتم‌های طبقه‌بند	[۷]	داده بانکی شبیه سازی شده	شبکه عصبی هم‌آمیختی + حافظه طولانی کوتاه مدتk + -نزدیک‌ترین همسایه	متعادل‌سازی	آخرین لایه تمام متصل
خودرمزگذار + الگوریتم‌های طبقه‌بند	[۱۲]	داده‌های اروپایی	خودرمزگذار+ تقویت گرادیان سبک وزن	پاک‌سازی داده‌ها+ متعادل‌سازی+ نرمال‌سازی+حذف ویژگی	لایه فشردهٔ خودرمزگذار
	[۵]	داده‌های اروپایی	خودرمزگذار+ جنگل تصادفی	متعادل‌سازی	لایه فشرده‌ی خودرمزگذار
خودرمزگذار + شبکه عصبی	[۱۳]	داده آلمانی	خودرمزگذار+ شبکه عصبی	نرمال‌سازی+ متعادل‌سازی	لایه فشرده‌ی خودرمزگذار

استخراج‌شده آموزش دیده‌اند. نتایج نشان دادند که در برخی معیارهای ارزیابی، عملکرد مدل‌ها به نزدیک به ۱۰۰ درصد رسید.

در [۱۲]، از خودرمزگذار همراه با الگوریتم تقویت گرادیان سبک وزن استفاده شده است. در این روش، ابتدا از خودرمزگذار برای کاهش ابعاد داده‌ها و استخراج ویژگی‌های کلیدی استفاده می‌شود و سپس با استفاده از الگوریتم تقویت گرادیان سبک وزن طبقه‌بندی انجام می‌شود. نتایج نشان می‌دهد که این روش به‌طور طبیعی با داده‌های نامتوازن مطابقت دارد و در مقایسه با روش‌های سنتی مانند الگوریتم k-نزدیک‌ترین همسایه ، جنگل تصادفی و الگوریتم تقویت گرادیان سبک وزن عملکرد بهتری ارائه می‌دهد.

در [۱۳]، ویژگی‌های معنادار و مهم توسط خودرمزگذار استخراج شده و این ویژگی‌ها به شبکه‌های عصبی دیگر برای طبقه‌بندی و شناسایی تراکنش‌های تقلبی داده شده‌اند. این مدل ترکیبی نه تنها نسبت به مدل‌های منفرد شبکه عصبی عملکرد بهتری داشته، بلکه در مقایسه با روش تحلیل مولفه‌های اصلی نیز نمایش بهتری از داده‌ها و کاهش بعد مؤثرتری ارائه کرده است. این نتایج نشان‌دهنده کارآمدی شبکه‌های عصبی عمیق در بهبود عملکرد روش‌های تشخیص تقلب کارت اعتباری است.

در بسیاری از مطالعات پیشین، روش‌های ترکیبی مختلفی برای شناسایی تقلب ارائه شده است. با این وجود، این پژوهش‌ها به گزارش نتایج عددی بسنده کرده‌اند و کمتر به تحلیل آماری پرداخته‌اند. در جدول۱ مروری مقایسه‌ای

از پژوهش‌های منتخب سال‌های ۲۰۲۰ تا ۲۰۲۵ ارائه شده است. در این جدول، مجموعه داده، مدل، پیش‌پرازش و لایه استخراج ویژگی ذکر شده است.

تحلیل منابع بررسی شده نشان می‌دهد که علی‌رغم عملکرد قابل قبول روش‌های ترکیبی مبتنی بر استخراج ویژگی از شبکه‌های عصبی و استفاده از الگوریتم‌های کلاسیک یادگیری ماشین، اغلب مقایسه‌ها فقط به ارزیابی عددی مدل‌ها بسنده کرده‌اند و آزمون‌های آماری برای بررسی پایداری و قابل‌اعتماد بودن نتایج، ارائه نشده است

در زمینه استخراج ویژگی از شبکه‌های عصبی نیز، در ادبیات موجود معیار دقیق و مشخصی برای تعیین لایه بهینه ارائه نشده است. به‌ویژه در معماری‌هایی که کمتر مورد استفاده قرار گرفته‌اند، مشخص نیست که کدام لایه می‌تواند بهترین بازنمایی را برای تشخیص تقلب فراهم کند. این خلأ می‌تواند منجر به استخراج ویژگی هایی با قدرت بازنمایی ناکافی شود، که در نهایت باعث کاهش دقت و پایداری مدل در مسائل حساس نظیر تشخیص تقلب می‌شود.

موضوع تفسیرپذیری مدل‌ها اغلب نادیده گرفته شده و کمتر مطالعه‌ای به بررسی میزان درک‌پذیری خروجی مدل از نظر کاربران نهایی یا کارشناسان بانکی پرداخته است.

این کاستی بیانگر شکاف‌های پژوهشی مهمی هستند که پژوهش حاضر در راستای پر کردن آنها طراحی شده است.

در ادامه، روش پیشنهادی ما که بر انتخاب لایه بهینه از شبکه عصبی هم‌آمیختی و پرسپترون چندلایه برای استخراج ویژگی، استفاده از الگوریتم تقویت گرادیان سبک وزن برای طبقه‌بندی با ویژگی‌های استخراج شده، انجام آزمون آماری و تقویت تفسیرپذیری مبتنی است، معرفی می‌شود.

۳- مجموعه داده و مدل‌های پایه

این بخش، مجموعه‌داده و پیش‌پردازش آن، معیارهای ارزیابی عملکرد مدل‌ها و ساختار مدل‌های پایه، شامل شبکه‌های عصبی و الگوریتم تقویت گرادیان سبک وزن را شرح می‌دهد.

۳-۱ مجموعه داده و پیش‌پردازش

مجموعه‌دادهٔ مورد استفاده در این مقاله شامل ۲۸٤۸۰۷ تراکنش کارت اعتباری اروپایی در سپتامبر ۲۰۱۳ و دارای دو کلاس «تقلب» و «عدم تقلب» است که از این میان تنها ٤۹۲ تراکنش تقلبی بوده و در نتیجه مجموعه‌داده به‌شدت نامتعادل است. این داده‌ها شامل ۳۰ ویژگی هستند، ۲۸ ویژگی ناشناس (که از V1 تا V28 نام‌گذاری شده‌اند) حاصل از تحلیل مؤلفه‌های اصلی و دو ویژگی اصلی «زمان» و «مقدار تراکنش» بدون تغییر هستند.

در مرحلهٔ پیش‌پردازش، داده‌های تکراری حذف و ویژگیِ زمان کنار گذاشته شد. این ویژگی تنها فاصله هر تراکنش از آغاز فرایند ثبت داده‌ها را نشان می‌دهد و زمان واقعی وقوع تراکنش‌ها را در طول روز یا هفته مشخص نمی‌کند. افزون بر این، الگوریتم مورد استفاده وابستگی‌های زمانی میان تراکنش‌ها را در نظر نمی‌گیرد.

برای ساخت مدل، داده‌ها به نسبت ۸۰ درصد برای آموزش (۲۵ درصد از آن برای اعتبارسنجی) و ۲۰ درصد برای آزمون نهایی تقسیم شدند. تقسیم داده‌ها پیش از هرگونه عملیات پیش‌پردازش انجام شد تا از بروز هرگونه نشت داده جلوگیری گردد.

برای نرمال‌سازی داده‌ها از روش Z-Score استفاده شد، به‌طوری‌که میانگین و انحراف معیار صرفاً بر اساس داده‌های آموزشی محاسبه گردید و همان پارامترها بدون هیچ‌گونه بازآموزی بر داده‌های اعتبارسنجی و آزمون اعمال شد. این رویکرد تضمین می‌کند که هیچ اطلاعاتی از توزیع داده‌های آزمون وارد فرآیند آموزش نشود.

به‌منظور رفع عدم‌تعادل کلاس‌ها، عملیات نمونه‌برداری افزایشی تنها بر روی داده‌های آموزشی نرمال‌شده اجرا شد. در این روش، با درون‌یابی میان هر نمونه از کلاس

اقلیت و K همسایه نزدیک آن، داده‌های مصنوعی جدید ایجاد می‌شود تا تعادل نسبی میان کلاس‌ها برقرار گردد. تعداد همسایه‌ها k=5 در نظر گرفته شد تا تنوع کافی در داده‌های تولید شده ایجاد شود و از هم‌پوشانی بیش از حد میان کلاس‌ها جلوگیری گردد. لازم به ذکر است که هیچ‌گونه نمونه‌برداری یا تغییر ساختاری بر روی مجموعه‌های اعتبارسنجی و آزمون اعمال نشد تا ساختار واقعی داده‌ها حفظ شود و ارزیابی مدل به صورت بی‌طرفانه انجام گیرد

۳-۲ معیارهای ارزیابی

معیارهای ارزیابی این پژوهش شامل صحت[18]، دقت[19]، بازیابی[20] و امتیاز 1F [21] هستند.

- صحت: نسبت کل پیش‌بینی‌های درست انجام شده به کل داده‌ها.
- دقت: نسبت پیش‌بینی‌های درست مثبت به کل پیش‌بینی‌های مثبت.
- بازیابی: نسبت داده‌های مثبت واقعی که درست شناسایی شده‌اند به کل داده‌های مثبت واقعی.
- امتیاز 1F: میانگین هارمونیک دقت و بازیابی.

و همچنین از دو مفهوم خطای مثبت کاذب و خطای منفی کاذب نیز استفاده شده است.

- خطای مثبت کاذب: مدل تراکنش «عدم تقلب» را به اشتباه «تقلب» تشخیص می‌دهد.
- خطای منفی کاذب: مدل تراکنش «تقلب» را به اشتباه «عدم تقلب» تشخیص می‌دهد.

۳-۳ ساختار مدل‌ها

در این بخش، شبکه‌های عصبی و الگوریتم تقویت گرادیان سبک وزن مورد استفاده، همراه با ویژگی‌ها و ساختار آنها ارائه شده‌است.

۳-۳-۱ ساختار شبکه‌های عصبی

رویکردهای مبتنی بر یادگیری عمیق، روش‌های پیشرفته‌ای هستند که برای تحلیل و شناسایی الگوهای پیچیده در مجموعه‌های داده بزرگ، مانند تراکنش‌های کارت اعتباری، استفاده می‌شوند[۱۴]. شبکه عصبی پرسپترون چندلایه و شبکه عصبی هم‌آمیختی از زیرشاخه‌های یادگیری عمیق هستند.

شبکه عصبی پرسپترون چندلایه از مجموعه‌ای از نورون‌ها تشکیل شده و این نورون‌ها از طریق وزن‌های پیوندی[22] به یکدیگر متصل‌اند. عملکرد پرسپترون چندلایه به‌گونه‌ای است که یک مجموعه ورودی را به خروجی دلخواه تبدیل می‌کند[۱۵]. شبکه عصبی پیچشی بر پایه ساختار پرسپترون چندلایه توسعه یافته‌است[۱۶].

شبکه‌های عصبی مورد استفاده در این پژوهش شامل دو مدل مستقل، پرسپترون چندلایه با چهار لایه و شبکه عصبی هم‌آمیختی با ۵ لایه بودند که آموزش آنها با حداکثر ۱۵۰ و ۳۰ دوره با استفاده از توقف زودهنگام برای جلوگیری از بیش‌برازش انجام شد و مدل‌ها پس از ۷۴ دوره و ۱۴ دوره به دلیل عدم پیشرفت معیار صحت در داده‌های اعتبارسنجی طی چهار دوره پیاپی، متوقف شدند

برای بهبود فرایند یادگیری مدل‌ها، از لایهٔ نرمال‌سازی دسته‌ای[23]، برای جلوگیری از بیش‌برازش از لایهٔ حذف تصادفی[24]، برای کاهش ابعاد ویژگی از لایه‌ی تجمیع بیشینه[25] و برای ایجاد توانایی یادگیری روابط غیرخطی در شبکه از توابع فعال‌ساز[26] LeakyReLU، ReLU و Sigmoid استفاده شد. در پایان آموزش، مدل پرسپترون چندلایه به صحتی ۰/۹۹۵۲ در داده‌های آزمون و ۰/۹۹۶۲ در داده‌های اعتبارسنجی دست یافت و مقادیر زیان متناظر به‌ترتیب برابر با ۰/۰۲۲ و ۰/۰۲۰ گزارش شد. همچنین، مدل شبکهٔ عصبی هم‌آمیختی صحتی معادل ۰/۹۹۹۹ در داده‌های آموزشی و ۰/۹۹۹۴ در داده‌های اعتبارسنجی

18-Accuracy
19-Precision
20-Recall
21-F1 Score

22-Linking Weights
23-Batch Normalization
24-Dropout
25-MaxPooling
26-Activation Functions

با مقادیر زیان ۰/۰۰۰۸۶ و ۰/۰۰۵۸ به‌دست آورد. نتایج مذکور بیانگر همگرایی مناسب هر دو مدل و عدم بروز بیش‌برازش محسوس در فرایند یادگیری هستند.

۳-۳-۲ الگوریتم تقویت گرادیان سبک وزن

ماشین تقویت گرادیان سبک‌وزن یک چارچوب قدرتمند برای پیاده‌سازی الگوریتم درخت تصمیم تقویت گرادیانی است که هم کارآمد بوده و هم به‌طور گسترده در پردازش مجموعه‌داده‌های بزرگ و ساختاریافته، به‌ویژه در حوزه‌های مالی مانند تشخیص تقلب کارت اعتباری، مورد استفاده قرار می‌گیرد. کارایی این الگوریتم در مدیریت مجموعه‌داده‌های بزرگ برای پردازش حجم بالای داده‌های تراکنش بسیار حیاتی است[۱۷].

بر خلاف بسیاری از ابزارهای سنتی درخت تصمیم تقویت گرادیانی که از رشد درخت بر پایه سطح[۲۷] استفاده می‌کنند، ماشین تقویت گرادیان سبک‌وزن از یک راهبرد رشد مبتنی بر برگ[۲۸] با محدودیت عمق بهره می‌برد. در این روش، در هر مرحله، برگی انتخاب می‌شود که بیشترین سود تقسیم را دارد، سپس آن برگ تقسیم می‌شود. این فرایند تا رسیدن به شرایط توقف ادامه می‌یابد. مزیت اصلی این رویکرد در مقایسه با روش سطحی آن است که می‌تواند با تعداد تقسیمات کمتر، خطای آموزش را کاهش داده و به دقت بالاتری برسد[۱۲].

۴- روش‌شناسی

در این بخش، معیارهای کمی به‌کاررفته برای انتخاب لایهٔ بهینه در استخراج ویژگی‌ها تشریح شده‌است، سپس فرایند استخراج ویژگی، آزمون آماری و تحلیل تفسیرپذیری مبتنی بر چارچوب SHAP توضیح داده می‌شود.

۴-۱ انتخاب لایه بهینه

در بسیاری از پژوهش‌های پیشین، فرایند انتخاب لایهٔ مناسب برای استخراج ویژگی‌ها از شبکه‌های عصبی به‌صورت شفاف توضیح داده نشده است. در اغلب موارد، یا اساس انتخاب لایه ذکر نمی‌شود، یا صرفاً بر پایهٔ نتایج تجربی، لایهٔ بهتر انتخاب می‌شود. این روش‌ها، علاوه بر ایجاد ابهام، باعث افزایش پیچیدگی محاسباتی و افزایش زمان اجرای مدل خواهند شد.

در این پژوهش، مجموعه‌ای از شاخص‌های کمّی معرفی و ارزیابی شده‌اند تا با تحلیل یکپارچه آنها بتوان بدون نیاز به آزمون‌های مکرر و پرهزینه، لایه‌ی بهینه برای استخراج ویژگی‌ها را انتخاب نمود. این شاخص‌ها عبارتند از:

۱) شاخص کالینسکی- هاراباز

ویژگی‌های استخراج‌شده باید بتوانند داده‌ها را به‌خوبی از یکدیگر متمایز کنند و شرایط مناسبی برای خوشه‌بندی فراهم نمایند بنابرین از شاخص کالینسکی- هاراباز[۲۹] که به‌صورت نسبت پراکندگی بین‌خوشه‌ای به پراکندگی درون‌خوشه‌ای تعریف می‌شود، برای ارزیابی کیفیت خوشه‌بندی استفاده شده است. مقدار بالاتر این شاخص نشان‌دهنده کیفیت بهتر خوشه‌بندی و در نتیجه، مناسب‌تر بودن ویژگی‌های استخراج‌شده برای تمایز داده‌ها است[۱۸]

با این وجود، شاخص کالینسکی- هاراباز به طوری ذاتی برای خوشه‌های دارای ساختار محدب طراحی شده است و در شرایطی که خوشه‌ها دارای ساختارهای پیچیده‌تر یا غیرمحدب هستند، ممکن است عملکرد این شاخص کاهش یابد[۱۸]، بنابراین بهتر است که از شاخص دیویس-بولدین[۳۰] به عنوان مکمل استفاده شده‌است.

۲) شاخص دیویس-بولدین

این شاخص میانگین شباهت بین هر خوشه و مشابه‌ترین خوشه دیگر را اندازه‌گیری می‌کند. این شباهت بر اساس نسبت پراکندگی درونی خوشه‌ها به فاصله بین مراکز آنها سنجیده می‌شود. هر چه مقدار این شاخص به صفر نزدیک‌تر باشد، نشان‌دهنده خوشه‌بندی بهتر و تفکیک‌پذیری بالاتر داده‌هاست[۱۸].

27-Level-wise
28-Leaf-wise
29-Calinski-Harabasz Index
30-Davies-Bouldin

۳) اطلاعات متقابل

اطلاعات متقابل[۳۱] مفهومی از نظریه اطلاعات است که میزان وابستگی بین جفت متغیرها را برآورد می‌کند. این معیار، بدون وابستگی به بازپارامترسازی، میزان وابستگی میان متغیرها را اندازه‌گیری می‌کند[۱۹]. هرچه این معیار بیشتر باشد، ارتباط آماری بین این دو متغیر قوی‌تر است

در این پژوهش، اطلاعات متقابل میان هر ویژگی و برچسب با استفاده از تخمین‌گر غیرپارامتریک مبتنی بر k–نزدیک‌ترین همسایه[۳۲] محاسبه شد. این روش وابستگی مستقیم به بُعد داده‌ها ندارد[۲۰] و می‌تواند روابط خطی و غیرخطی میان ویژگی‌ها و برچسب را شناسایی کند[۲۱].

٤) افزونگی

یک ویژگی بولی[۳۳]، افزونه[۳٤] است اگر حذف آن تاثیری بر هیچ یک از جفت نمونه‌هایی که در سایر ویژگی‌ها کاملاً یکسان هستند، نداشته باشد.

برای محاسبه افزونگی، از مفهوم اطلاعات متقابل استفاده می‌شود. در حالی که اطلاعات متقابل معرفی‌شده در بخش قبل، میزان همبستگی بین ویژگی‌ها و برچسب‌ها را اندازه‌گیری می‌کند، می‌توان از همان روابط ریاضی برای سنجش همبستگی بین ویژگی‌ها با یکدیگر نیز بهره گرفت.

اگر مقدار اطلاعات متقابل بین دو ویژگی زیاد باشد، این نشان‌دهنده افزونگی اطلاعاتی بین آن‌هاست؛ به این معنی که یکی از آنها احتمالاً اطلاعات جدیدی نسبت به دیگری ارائه نمی‌دهد و وجود آن ممکن است تنها موجب افزایش حجم داده و اشغال حافظه شود.

۵) ابعاد لایه‌ها

در این پژوهش، ابعاد خروجی لایه‌ها به‌منظور تحلیل پیچیدگی محاسباتی مورد بررسی قرار گرفته‌اند.

در ادبیات مرتبط با این حوزه، هیچ روش استاندارد یا اجماعی برای تعیین وزن نسبی میان این معیارها گزارش نشده است. بنابراین، برای جلوگیری از ایجاد سوگیری و حفظ شفافیت، وزن برابر برای تمام معیارها در نظر گرفته شد. در گام نهایی، کلیه معیارهای ارزیابی مربوط به هر لایه با استفاده از روش حداقل- حداکثر[۳۵] نرمال شدند. در این فرایند، به این نکته توجه شده است که در برخی معیارها، مقدار کمتر نشان‌دهنده عملکرد بهتر است، در حالی که در برخی دیگر، مقدار بیشتر مطلوب‌تر است، بنابراین نرمال‌سازی مطابق با ماهیت هر معیار انجام شده است. در نهایت، میانگین نهایی تمام معیارهای استانداردشده برای هر لایه محاسبه گردیده و لایه‌ای که بالاترین میانگین را به‌دست آورده، به‌عنوان لایه بهینه انتخاب و مورد استفاده قرار گرفته است.

۴-۲ استخراج ویژگی

پس از انتخاب لایهٔ بهینه، ویژگی‌های استخراج‌شده با تکنیک Z-score، نرمال‌سازی و به الگوریتم تقویت گرادیان سبک‌وزن منتقل شدند.

به‌منظور اطمینان از پایداری عملکرد مدل و کاهش احتمال تصادفی بودن نتایج، فرایند آموزش و ارزیابی پنج مرتبه تکرار گردید. در هر تکرار، تقسیم‌بندی داده‌ها بین مجموعه‌های آموزش و آزمون ثابت بود و تنها حالت تصادفی شروع آموزش الگوریتم طبقه‌بند تغییر می‌کرد. در هر تکرار، مدل با استفاده از شاخص‌های صحت، دقت، بازیابی، امتیاز F1، خطای مثبت کاذب و خطای منفی کاذب مورد ارزیابی قرار گرفت. در پایان، برای هر یک از معیارهای مذکور به‌صورت مستقل، میانگین مقادیر به‌دست‌آمده در پنج اجرا محاسبه و به‌عنوان نتیجه نهایی گزارش شد.

۴-۳ آزمون آماری

به منظور ارائه مقایسه عددی دقیق‌تر و تقویت تحلیل آماری، مدل منفرد الگوریتم سبک وزن نیز ارزیابی شد. مشابه مدل‌های ترکیبی، این مدل پنج بار اجرا شد و

31-Mutual Information
32-k-Nearest Neighbors non-parametric estimator
33-Boolean Feature
34-Redundancy
35-Min-Max Normalization

میانگین نتایج برای مقایسه با دو مدل دیگر مورد استفاده قرار گرفت تا اثر تصادفی بودن نتایج به حداقل برسد. برای ارزیابی معناداری آماری روش پیشنهادی، آزمون ویلکاکسون[36] برای سه گروه مختلف شامل مدل ترکیبی مبتنی بر شبکه عصبی پیچشی، مدل ترکیبی مبتنی بر شبکه عصبی پرسپترون چندلایه و مدل منفرد انجام شد. مقایسه‌ها به‌صورت جفتی و با در نظر گرفتن تمام معیارهای ارزیابی صورت گرفت.

۴-۴ تفسیرپذیری

هدف از به‌کارگیری تفسیرپذیری در این پژوهش، شفاف‌سازی سازوکار تصمیم‌گیری مدل و مشخص‌کردن این موضوع است که کدام ویژگی‌های استخراج‌شده بیشترین نقش را در طبقه‌بندی تراکنش‌ها داشته‌اند. از آنجا که سامانه‌های تشخیص تقلب در محیط‌های حساس مالی مورد استفاده قرار می‌گیرند، امکان توضیح و توجیه تصمیم مدل برای تحلیل‌گران و واحدهای کنترلی اهمیت بالایی دارد. تفسیرپذیری کمک می‌کند تا در صورت شناسایی یک تراکنش به‌عنوان تقلب یا عدم تقلب، مشخص شود کدام متغیرهای اولیه بیشترین تأثیر را در فعال‌شدن ویژگی‌های کلیدی مدل داشته‌اند. این موضوع علاوه‌بر افزایش شفافیت و اعتمادپذیری مدل، امکان بازبینی هدفمند هشدارها و تنظیم مناسب‌تر آستانه‌های تصمیم‌گیری را فراهم می‌سازد. با هدف بررسی تفسیرپذیری تصمیم مدل، تحلیل مبتنی بر SHAP [37] بر روی مدلی که عملکرد بهتری نشان داده بود، انجام شد. در این تحلیل، علاوه بر شناسایی ویژگی‌های مؤثر استخراج‌شده توسط شبکه عصبی که در تصمیم نهایی الگوریتم طبقه‌بند نقش دارند، میزان تأثیر هر یک از این ویژگی‌ها و همچنین ویژگی‌های اولیه داده‌ها که شبکه عصبی برای ساخت مؤثرترین ویژگی ترکیب کرده بود، مورد بررسی قرار گرفت.

36-Wilcoxon signed-rank test
37-SHapley Additive exPlanations

۵- نتایج

در این بخش، نتایج حاصل از مراحل مختلف پژوهش شامل انتخاب لایه بهینه برای استخراج ویژگی، ارزیابی عملکرد مدل‌ها، بررسی معناداری تفاوت میان آن‌ها و تحلیل تفسیرپذیری مدل منتخب ارائه می‌شود .هدف از این بخش، ارزیابی اثربخشی رویکردهای پیشنهادی در ارتقای عملکرد تشخیص تقلب و تحلیل نقش ویژگی‌های استخراج‌شده در فرآیند تصمیم‌گیری مدل است.

۵-۱ نتیجه انتخاب بهینه لایه

برای انتخاب لایه بهینه جهت استخراج ویژگی، مجموعه‌ای از معیارهای کمّی شامل شاخص کالینسکی- هاراباز، شاخص دیویس-بولدین، اطلاعات متقابل، میزان افزونگی ویژگی‌ها و ابعاد لایه بررسی شد. با توجه به این که برخی معیارها (کالینسکی- هاراباز و اطلاعات متقابل) مقادیر بالاتر و برخی دیگر (ابعاد لایه، دیویس- بولدین و افزونگی) مقادیر کمتر، نشان‌دهنده کیفیت بهتر هستند، پیش از اعمال تکنیک نرمال‌سازی حداقل- حداکثر جهت هر شاخص به‌گونه‌ای تنظیم شد که قابلیت مقایسه عادلانه فراهم گردد. مقادیر نرمال‌شده در جدول ۲ و میانگین آنها برای تصمیم‌گیری نهایی در جدول ۳ ارائه شده است.

بر اساس نتایج به‌دست آمده، لایه تمام متصل از شبکه عصبی هم‌آمیختی و لایه پنهانِ تمام‌متصل سوم از شبکه عصبی پرسپترون چندلایه به‌عنوان لایه‌های بهینه شناسایی شدند. ویژگی‌های استخراج‌شده از این لایه‌ها بعد از نرمال‌سازی به الگوریتم تقویت گرادیان سبک‌وزن تزریق شد.

۵-۲ نتایج مدل‌ها

همان‌طور که در بخش ۲-٤ و ۳-٤ توضیح داده شد، هر سه مدل شامل مدل ترکیبی مبتنی بر شبکه عصبی هم‌آمیختی، مدل ترکیبی مبتنی بر پرسپترون چندلایه و مدل منفرد، پنج مرتبه اجرا شدند. مقادیر معیارهای عملکرد

جدول ۲. مقادیر نرمال‌شده معیارهای کمی ارزیابی لایه‌های شبکه عصبی

نام شبکه عصبی	لایه‌های شبکه عصبی هم‌آمیختی					لایه‌های شبکه عصبی پرسپترون چندلایه			
نام لایه / معیار	لایه گسترش‌یافته تمام متصل	لایه هم‌آمیختی یک بعدی اول	لایه هم‌آمیختی یک بعدی دوم	لایه تخت‌سازی	لایه‌های تمام متصل	لایه پنهانِ تمام‌متصل اول	لایه پنهانِ تمام‌متصل دوم	لایه پنهانِ تمام‌متصل سوم	لایه پنهانِ تمام‌متصل چهارم
میزان افزونگی بین ویژگی‌ها	۰/۹۶	۱	۰/۸۲	۰/۵	۰	۱	۰/۵۴	۰	۰/۴۶
شاخص دیویس- بولدین	۰/۰۹۷	۰/۰۰۰۲۲	۰/۴۸۱	۰/۷۲۵	۱	۰	۰/۷۷	۱	۰/۷۳
شاخص کالینسکی- هارباز	۰/۰۰۱	۰	۰/۰۴	۰/۱۲	۱	۰	۰/۵۹	۱	۰/۴۳
اطلاعات متقابل	۰/۰۵	۰	۰/۴۷	۱	۰/۹	۰	۰/۶۲	۰/۹۹	۱
ابعاد ویژگی هر لایه	۰	۱	۱	۰	۱	۰	۰/۵۷	۰/۸۵	۱

جدول ۳. میانگین مقادیر نرمال‌شده معیارها برای انتخاب لایه بهینه

نام شبکه	نام لایه	نمره نهایی
شبکه عصبی هم‌آمیختی	لایه گسترش‌یافته تمام متصل	۰/۲۲
	لایه هم‌آمیختی یک بعدی اول	۰/۴
	لایه هم‌آمیختی یک بعدی دوم	۰/۵۶
	لایه تخت‌سازی	۰/۴۷
	لایه تمام متصل	۰/۷۸
شبکه عصبی پرسپترون چندلایه	لایه پنهانِ تمام متصل اول	۰/۲
	لایه پنهانِ تمام متصل دوم	۰/۶۲
	لایه پنهانِ تمام متصل سوم	۰/۷۷
	لایه پنهانِ تمام متصل چهارم	۰/۷۶

هر اجرا ثبت و میانگین نتایج برای مقایسه نهایی محاسبه شد و در جدول ٤ گزارش شده است.

بررسی نتایج نشان می‌دهد که استفاده از ویژگی‌های استخراج‌شده از شبکه عصبی هم‌آمیختی در مقایسه با ویژگی‌های استخراج‌شده از شبکه عصبی پرسپترون چندلایه و مدل منفرد با ویژگی‌های اولیه، منجر به بهبود عملکرد مدل شده است. مدل ترکیبی شبکه عصبی هم‌آمیختی + الگوریتم تقویت گرادیان سبک وزن در اغلب معیارها برتری محسوسی دارد. به‌طور کلی، ترکیب عصبی هم‌آمیختی با الگوریتم تقویت گرادیان سبک وزن توانسته است تعادل بهتری میان معیارهای اصلی عملکرد برقرار سازد اما هیچ کدام از مدل‌های ترکیبی، با وجود

جدول ۴. نتایج مدل‌های ترکیبی و منفرد بر اساس معیارهای عملکرد

نام مدل	صحت	دقت	بازیابی	امتیاز f1	خطای مثبت کاذب	خطای منفی کاذب
شبکه عصبی هم‌آمیختی + الگوریتم تقویت گرادیان سبک وزن	۹۹/۹۵	۷۵/۵۵	۹۱/۶۴	۸۲/۸۲	۶	۲۲
شبکه عصبی پرسپترون چندلایه + الگوریتم تقویت گرادیان سبک وزن	۹۹/۹۳	۷۱/۱۱	۸۲/۴۹	۷۶/۳۸	۱۴	۲۶
الگوریتم تقویت گرادیان سبک وزن	۹۹/۸۳	۵۷/۸۳	۷۷/۷۷	۶۵/۷۹	۵۳	۲۱

بهبود عملکرد، در کاهش خطای منفی کاذب موفق نبوده و با افزایشی اندک همراه بوده است. این رفتار عمدتاً ناشی از ساختار ترکیبی شبکه‌های عصبی و الگوریتم تقویت گرادیان سبک‌وزن است. در مرحله استخراج ویژگی، بخش عصبی الگوهای پرتکرار و عمومی داده را در قالب نمایش‌های فشرده‌تر بازنمایی می‌کند که اگرچه موجب افزایش دقت کلی مدل می‌شود، اما می‌تواند موجب کاهش تمایزهای ظریف میان تراکنش‌های «تقلب» و «عدم تقلب» گردد. از سوی دیگر، الگوریتم تقویت گرادیان سبک‌وزن،

بـــه دلیل ماهیـــت درخت‌گونه و راهبرد تقویـــت گرادیانی، تقسیم‌هـــای محافظه‌کارانـــه‌ای را در نواحی با بیشـــترین اطمینان انجـــام می‌دهد و بنابراین خطـــای مثبت کاذب را کاهش می‌دهد. به این ترتیب، میانگین خطای منفی کاذب در چند اجرای مستقل مدل ترکیبی تغییر قابل‌توجهی نداشته و این رفتار می‌تواند به‌عنوان یک ویژگی ساختاری و پایدار مدل تلقی شـــود. با این وجود، این افزایش در مقایســـه با کاهش محســـوس خطای نوع اول قابل قبول است و نشان می‌دهد که تعادل میان دقت و بازیابی حفظ شده است.

۵-۳ تحلیل آماری

بـــرای ارزیابی معناداری تفاوت با اســـتفاده از آزمون ویلکاکسون، مقایســـه جفتی بین مدل‌ها صورت گرفت، که نتایج آن در جدول ۵ گزارش شده است.

جدول ۵. مقایسه جفتی بین مدل‌ها با آزمون ویلکاکسون

مقایسه مدل‌ها	p-value	آماره آزمون
شبکه عصبی هم‌آمیختی + الگوریتم تقویت گرادیان سبک وزن و شبکه عصبی پرسپترون چندلایه + الگوریتم تقویت گرادیان سبک وزن	۰/۰۶۲۵	۰
شبکه عصبی هم‌آمیختی + الگوریتم تقویت گرادیان سبک وزن و الگوریتم تقویت گرادیان سبک وزن	۰/۰۶۲۵	۰
شبکه عصبی پرسپترون چندلایه + الگوریتم تقویت گرادیان سبک وزن و الگوریتم تقویت گرادیان سبک وزن	۰/۰۶۲۵	۰

نتایج نشان می‌دهد که مقدار آماره آزمون[38] در تمامی مقایســـه‌ها برابر با صفر اســـت که بدین معناســـت جهت اختلاف‌هـــا در تمامی موارد در یک‌راســـتا بوده و برتری قطعی میان مدل‌ها وجود ندارد. همچنین مقدار p-value برای هر سه مقایسه برابر با ۰/۰۶۲۵ گزارش شد که بزرگتر از ســـطح خطای متداول (۰/۰۵) است و می‌توان نتیجه گرفت که تفاوت میان مدل‌ها از نظر آماری معنادار نیست. با این حال، عدم معناداری آماری در این پژوهش به معنای فقدان برتری کاربردی نیست. بررسی شاخص‌های عملکرد نشان می‌دهد که مدل‌های پیشـــنهادی توانســـته‌اند میزان خطای مثبت کاذب را نسبت به مدل منفرد کاهش دهند. موضوعی که در سامانه‌های واقعی تشخیص تقلب اهمیت حیاتی دارد، زیرا کاهش هشدارهای اشتباه موجب جلوگیری از مسدود شـــدن غیرضـــروری کارت‌های بانکی مشـــتریان و حفظ تجربه کاربری و رضایت مشتریان می‌شود.

۵-۴ تفسیرپذیری

از میان سه مدل مقایسه‌شده، مدل ترکیبی شبکه عصبی هم‌آمیختی + الگوریتم تقویت گرادیان ســـبک وزن به دلیل عملکرد بهتر نسبت به ســـایر مدل‌ها به‌عنوان مدل منتخب جهت تحلیل تفسیرپذیری انتخاب شد. هدف از تفسیرپذیری، بررسی نحوه تأثیرگذاری ویژگی‌های استخراج‌شده توسط شبکهٔ عصبی هم‌آمیختی بر تصمیم‌گیری نهایی است. برای این منظور، از تحلیل مقادیـــر SHAP بر خروجی الگوریتم تقویت گرادیان ســـبک وزن اســـتفاده گردیـــد. از آنجا که خروجی شبکه شامل ویژگی‌های ترکیبی بدون نام مشخص بود، این ویژگی‌ها به‌ترتیب خروجی شـــماره‌گذاری شدند. در شکل ۱، بیســـت ویژگی با بیشترین میانگین قدر مطلق مقادیر SHAP ، به‌عنوان اثرگذارترین ویژگی‌ها نمایش داده شده‌اند. مطابق با شـــکل۱، ویژگی ۲، یعنی دومین ویژگی استخراج‌شـــده از شبکهٔ عصبی هم‌آمیختی، بیشترین سهم را در فرآیند تصمیم‌گیری مدل داشته است. برای بررسی دقیق‌تر این بیســـت ویژگی موثـــر، مقادیر SHAP برای هر نمونه داده محاســـبه و نتایج در شـــکل ۲ نشان ارائه شده اســـت. همان‌طور که در شکل ۲ مشـــاهده می‌شود، مقادیر SHAP در محور افقی نشـــان‌دهنده جهت و میزان تأثیر هر ویژگی بر خروجی مدل هستند؛ مقادیر مثبت بیانگر افزایش احتمـــال وقوع «تقلـــب» و مقادیر منفی بیانگـــر کاهش این احتمال‌اند. هر نقطه یک نمونه تراکنش را نشـــان می‌دهد و رنگ نقاط مقدار واقعی ویژگی را بازنمایی می‌کند. الگوهای مشـــاهده شده بیانگر آن اســـت که تنها تعداد محدودی از ویژگی‌ها دارای تأثیر مثبت یا منفی معنادار هستند و بخش عمـــده ویژگی‌ها نقشی محدود در تصمیم‌گیری مدل دارند. ایـــن موضوع حاکی از آن اســـت که مدل بـــرای تمایز بین

38-Statistic

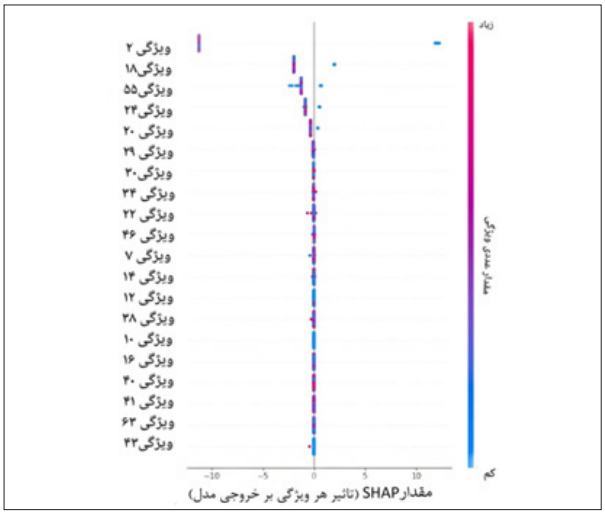
دو کلاسِ «تقلب» و «عدم تقلب» از مجموعه‌ای مشخص و تکرارشونده از عوامل استفاده می‌کند. به‌ویژه، ویژگی شماره ۲ بیشترین تأثیر را بر افزایش یا کاهش احتمال تقلب داشته و نقش اصلی را در تفکیک نمونه‌های مشکوک از تراکنش‌های عادی ایفا کرده است. از آنجا که ویژگی‌های مورد استفاده‌ی الگوریتم طبقه‌بند، ترکیبی خطی از ویژگی‌های اولیهٔ استخراج‌شده هستند، برای شناسایی مهمترین مؤلفه‌های تشکیل‌دهنده ویژگی ۲، یک مدل رگرسیون خطی آموزش داده شد. در این مدل، ویژگی‌های اولیه به‌عنوان متغیرهای مستقل و ویژگی میانی حاصل از شبکهٔ عصبی به‌عنوان متغیر وابسته در نظر گرفته شدند. نتایج این تحلیل در شکل ۳ نشان داده شده است که بر اساس آن، ویژگی «مقدار تراکنش» بیشترین سهم را در شکل‌گیری ویژگی ۲ دارد.

بنابراین، می‌توان گفت بخش قابل‌توجهی از اطلاعاتی که ویژگی ۲ در تصمیم‌گیری مدل منتقل می‌کند، به «مقدار تراکنش» مربوط است. با توجه به این که ویژگی ۲ مهمترین ویژگی در طبقه‌بندی نهایی شناخته شد، این نتیجه بیان می‌کند که اطلاعات مرتبط با مقدار تراکنش نقش محوری در عملکرد مدل تشخیص تقلب داشته است.

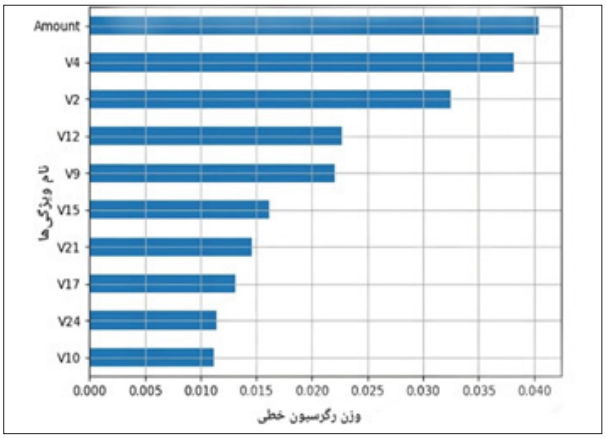
۵-۵ جمع‌بندی

در این پژوهش، سه گروه مدل شامل یک مدل منفرد مبتنی بر ویژگی‌های اولیه و دو مدل ترکیبی مبتنی بر استخراج ویژگی از شبکه‌های عصبی (شبکه عصبی هم‌آمیختی و پرسپترون چندلایه) و ترکیب آنها با الگوریتم تقویت گرادیان سبک‌وزن مورد بررسی قرار گرفتند. نتایج نشان داد که مدل‌های ترکیبی به‌طورکلی عملکرد بهتری نسبت به مدل منفرد دارند. همچنین، در میان مدل‌های ترکیبی، مدل مبتنی بر ویژگی‌های استخراج‌شده از شبکه عصبی هم‌آمیختی در اغلب معیارهای ارزیابی به‌ویژه در کاهش خطای مثبت کاذب، برتری بیشتری نسبت به مدل مبتنی بر ویژگی‌های استخراج شده از شبکه پرسپترون چندلایه نشان داد. از سوی دیگر، مقدار خطای منفی کاذب در مدل‌های ترکیبی با افزایشی اندک همراه بوده است. این

شکل ۱. بیست ویژگی موثر در تصمیم‌گیری مدل



شکل ۲. توزیع مقادیر SHAP برای بیست ویژگی موثر



شکل ۳. ده ویژگی اولیه مؤثر در شکل‌گیری ویژگی شماره ۲ از لایه تمام‌متصل شبکهٔ عصبی هم‌آمیختی

افزایش در مقایسه با کاهش قابل‌توجه خطای مثبت کاذب، از نظر عملیاتی قابل پذیرش است و نشان می‌دهد که مدل پیشنهادی توانسته است تعادل مؤثری میان دقت و بازیابی برقرار سازد.

همچنین، برای انتخاب لایه‌های مناسب جهت استخراج ویژگی از شبکه‌های عصبی، برای هر لایه مجموعه‌ای از معیارهای کمّی شامل شاخص کالینسکی- هاراباز، شاخص دیویس- بولدین، اطلاعات متقابل، میزان افزونگی ویژگی‌ها و ابعاد لایه محاسبه شد. پس از نرمال‌سازی این معیارها و یکسان‌سازی جهت آن‌ها، میانگین نمرات نرمال‌شده به‌عنوان معیار تصمیم‌گیری نهایی مورد استفاده قرار گرفت و لایه‌ای که بالاترین میانگین را کسب کرده بود، به‌عنوان لایهٔ مناسب برای استخراج ویژگی انتخاب شد. سپس، ویژگی‌های حاصل از لایه‌های منتخب پس از نرمال‌سازی مجدد، به الگوریتم تقویت گرادیان سبک‌وزن برای مرحلهٔ طبقه‌بندی منتقل شدند.

تحلیل تفسیرپذیری نشان داد که ویژگی شماره ۲، که از شبکهٔ عصبی هم‌آمیختی استخراج شده و بیشترین سهم را در تصمیم‌گیری مدل داشته است، عمدتاً تحت تأثیر متغیر «مقدار تراکنش» شکل گرفته است. به عبارت دیگر، مدل از مقدار تراکنش به‌عنوان ورودی پایه برای ساخت یک ویژگی ترکیبی مؤثر استفاده کرده است و این ویژگی نقش کلیدی در فرایند تشخیص تقلب داشته است. این نتیجه نشان می‌دهد که اطلاعات مربوط به مقدار تراکنش، پس از تبدیل و بازنمایی توسط شبکه عصبی، در بهبود عملکرد مدل طبقه‌بندی نقش قابل‌توجهی ایفا کرده است.

۶- نتیجه‌گیری

این بخش به جمع‌بندی نتایج پژوهش و ارائه راهکارهایی برای مطالعات آتی اختصاص دارد.

۶-۱ یافته‌های کلیدی پژوهش

در این پژوهش، رویکردی ترکیبی برای تشخیص تقلب در تراکنش‌های کارت اعتباری ارائه شد که در آن ویژگی‌های استخراج‌شده از شبکه‌های عصبی هم‌آمیختی و پرسپترون چندلایه به الگوریتم تقویت گرادیان سبک‌وزن منتقل شدند. برای انتخاب لایه مناسب برای استخراج ویژگی، از ترکیب چند معیار کمّی استفاده شد و لایه‌ای که بالاترین میانگین نرمال‌شده را کسب کرد، به‌عنوان لایهٔ بهینه تعیین گردید.

نتایج نشان داد که مدل‌های ترکیبی، به‌ویژه ترکیب مبتنی بر شبکه عصبی هم‌آمیختی، در مقایسه با مدل منفرد موجب بهبود امتیاز F1 و کاهش قابل‌توجه خطای مثبت کاذب شدند. در مقابل، مقدار خطای منفی کاذب با افزایشی اندک همراه بود، افزایشی که در مقایسه با کاهش قابل توجه خطای مثبت کاذب، توازن عملکرد مدل را تحت تأثیر قرار نمی‌دهد. آزمون ویلکاکسون نیز تفاوت میان مدل‌های ترکیبی و مدل منفرد را از نظر آماری معنادار ندانست. با این حال، کاهش خطای مثبت کاذب از دیدگاه عملی در سامانه‌های بانکی اهمیت دارد، زیرا منجر به کاهش مسدودسازی غیرضروری کارت‌های سالم و بهبود تجربه مشتریان می‌شود.

تحلیل تفسیرپذیری مدل بر پایه‌ی SHAP مشخص کرد که دومین ویژگی استخراج‌شده از شبکه عصبی هم‌آمیختی بیشترین نقش را در طبقه‌بندی داشته و این ویژگی عمدتاً بر پایه‌ی اطلاعات مرتبط با مقدار تراکنش شکل گرفته است. این موضوع نشان می‌دهد که شبکه عصبی توانسته است نمایش ساخت‌یافته‌تری از متغیرهای اولیه ایجاد کند و این نمایش جدید در بهبود عملکرد مدل تأثیرگذار بوده است.

در مجموع، یافته‌ها بیان می‌کنند که ترکیب استخراج ویژگی از شبکه‌های عصبی و مدل‌های تقویت گرادیان می‌تواند راهکاری مؤثر برای افزایش کیفیت سیستم‌های تشخیص تقلب بانکی باشد، به‌ویژه در شرایطی که کاهش خطاهای مثبت کاذب اهمیت عملی بالایی دارد.

۶-۲ پیشنهادهایی برای پژوهش‌های آینده

در ادامه، با توجه به یافته‌های این پژوهش، پیشنهادهایی

برای توسعه مطالعات آینده ارائه می‌شود.

۱.در ایـــن پژوهش، تمامی معیارهای کمّی برای تعیین لایه بهینه با وزن برابر در نظر گرفته شدند. در مطالعات آینده می‌توان اهمیت نســـبی هر معیـــار را به‌صورت جداگانه بررسی کرد تا ســـهم واقعی هر شاخص در تعیین لایه بهینه مشخص شود.

۲.به‌کارگیری روش‌های مهندسی ویژگی، مانند کاهش ابعاد داده‌ها یا تمرکز بر انتخاب ویژگی‌های موثر، می‌تواند به بهبود مدل و کاهش پیچیدگی محاسباتی کمک کند.

۳.از معماری‌های متنوع شبکه‌های عصبی، مانند شبکه‌های ترنسفورمرها[39]، برای استخراج ویژگی‌ها استفاده شود.

٤.پیشنهاد می‌شود روش ارائه‌شده بر روی مجموعه‌داده‌های گوناگون و به‌روزتر (برای مثال داده‌های تراکنش برخط یا تراکنش‌های مبتنی بر رمز دوم پویا) نیز آزموده شود تا میـــزان تعمیم‌پذیری و پایداری مدل در ســـناریوهای واقعی مشخص گردد.

منابع

1. Taha, A. A. & Malebary S. J. 2020 An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine, IEEE Access, Vol. 8, pp. 25579–25587.
2. Carneiro, N., Figueira, G., & Costa, M. 2017. A data mining based system for credit-card fraud detection in e-tail, Decision Support Systems, Vol. 95, pp. 91–101.
3. Umaru, I. A., Aliyu, A. A., Ibrahim, M., Abdulkadir, S., Ahmed, M. A., Abubakar, M. A., & Tanko, S. A. 2025. An enhanced hybrid model combining LSTM, ResNet, and an attention mechanism for credit card fraud detection, FUDMA Journal of Science, Vol. 9, No. 2, pp. 42–48.
4. Carcillo, F., Le Borgne, Y.-A., & Bontempi, G. 2018. Streaming active learning strategies for real-life credit card fraud detection, International Journal of Data Science and Analytics, Vol. 7, No. 1, pp. 33–45.
5. Lin, T. H. & Jiang, J. R. 2021. Credit card fraud detection with autoencoder and probabilistic random forest, Mathematics, Vol. 9, No. 21, pp. 16-1.
6. Berhane, T., Melese, T., Walelign, A., & Mohammed, A. 2023. A hybrid convolutional neural network and support vector machine-based credit card fraud detection model, Mathematical Problems in Engineering, Vol. 2023, No. 1, pp. 1–10.
7. Nancy, A. M., Kumar, G. S., Veena, S., Vinoth, N. S., & Bandyopadhyay, M. 2020. Fraud detection in credit card transaction using hybrid model, AIP Conference Proceedings, pp. 1–10.
8. Singh, S., Nimje, H., Fulpatile, A., & Neware, R. 2024. Credit card fraud detection using a hybrid machine learning algorithm, Preprints.
9. Boukhalfa, K., Harous, S., & Bouras, A. 2022. A novel text2IMG mechanism of credit card fraud detection: A deep learning approach, Electronics, Vol. 11, No. 4, p. 756.
10. Subagio, A. & Utama, D. N. 2025. Fraud credit card transaction detection using hybrid multilayer perceptron-random forest method, Edelweiss Applied Science and Technology, Vol. 9, No. 3, pp. 2482–2494.
11. Ming, R., Abdelrahman, O., Innab, N., & Ibrahim, M. H. K. 2024. Enhancing fraud detection in auto insurance and credit card transactions: A novel approach integrating CNNs and machine learning algorithms, PeerJ Computer Science, Vol. 10, p. e2088.
12. Du, H., Lv, L., Guo, A., & Wang, H. 2023. AutoEncoder and LightGBM for credit card fraud detection problems, Symmetry, Vol. 15, No. 4, p. 870.
13. 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, Vol. 217, p. 119562.
14. Mienye, I. D. & Jere, N. 2024. Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions, IEEE Access, Vol. 12, pp. 96893–96910.
15. Kasasbeh, B., Aldabaybah, B. & Ahmad, H. 2022. Multilayer perceptron artificial neural networks-based model for credit card fraud detection, Indonesian Journal of Electrical Engineering and Computer Science, Vol. 26, No. 1, pp. 362–373.
16. Ke, Q., Liu, J., Bennamoun, M., An, S., Sohel, F., & Boussaid, F. 2018. Computer Vision for Human-Machine Interaction, in *Computer Vision for Assistive Healthcare, pp. 127–145.
17. Zhao, X., Liu, Y., & Zhao, Q.2024. Improved LightGBM for extremely imbalanced data and application to credit card fraud detection, IEEE Access, Vol. 12, pp. 159316–159335.
18. Gan و M. & Zhang, L. 2021. Iteratively local fisher score for feature selection, Applied Intelligence, Vol. 51, No. 8, pp. 6167–6181.
19. Han, W., Chen, H., & Poria, S. 2021. Improving multimodal fusion with hierarchical mutual information maximization for multimodal sentiment analysis, arXiv preprint arXiv:2109.00412.
20. Carrara, N. & Ernst, J. 2020. On the estimation of mutual information, Proceedings, vol. 33, no. 1, p. 31-39.
21. Runge, J. 2018. Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information, in Proceedings of the International Conference on Artificial Intelligence and Statistics, pp. 938–947.s

39-Transformers