

توسعه چارچوبی برای مدیریت ریسک‌های هوش مصنوعی در صنعت خدمات مالی

طاهره سجادی راد^۱

عبدالرسول مستاجران^۲

تاریخ دریافت: ۱۴۰۳/۰۸/۱۵ تاریخ چاپ: ۱۴۰۳/۱۱/۳۰

چکیده

در سال‌های اخیر، مؤسسات مالی به طور فزاینده‌ای از فناوری‌های پیشرفته از جمله هوش مصنوعی در مواردی مانند امتیازدهی اعتباری، شناسایی ناهنجاری‌ها، کنترل‌های داخلی و انطباق با مقررات استفاده می‌کنند که با دسترسی به مدل‌های یادگیری ماشین، استفاده از آن‌ها تسهیل شده است. با وجود دقت بالای پیش‌بینی این مدل‌ها، ممکن است با ریسک‌هایی همراه باشند؛ بنابراین، برای ذینفعان مربوطه، مانند کاربران تجاری، حساب‌رسان، ناظران و مشتریان نهایی قابل اعتماد نیستند. برای اندازه‌گیری قابلیت اعتماد برنامه‌های هوش مصنوعی، لازم است تا شاخص‌های کلیدی ریسک‌های هوش مصنوعی در صنعت خدمات مالی شناسایی شوند و بر اساس آن چارچوب مدیریت ریسک هوش مصنوعی ایجاد و الزامات نظارتی دقیقی تبیین شود. این مقاله، مروری بر برخی از ریسک‌های هوش مصنوعی در صنعت خدمات مالی داشته و تلاش کرده است تا چارچوبی برای مدیریت ریسک‌های هوش مصنوعی در این صنعت را معرفی کند. در این پژوهش پیشنهاد شده است که برای مدیریت ریسک هوش مصنوعی باید یک رویکرد یکپارچه اتخاذ شود که تمامی ریسک‌های اصلی مدل هوش مصنوعی را به طور همزمان در نظر می‌گیرد. به طور خاص، یک مدل یکپارچه مدیریت ریسک هوش مصنوعی که شامل سنجش مجموعه‌ای از چهار معیار پایداری، دقت، انصاف و توضیح‌پذیری با روش‌های آماری است، توسعه یابد. این پژوهش می‌تواند پایه‌گذار پژوهش‌های تجربی و کاربردی در زمینه اندازه‌گیری و مدیریت ریسک‌های هوش مصنوعی باشد.

واژگان کلیدی

مدل‌های یادگیری ماشین، هوش مصنوعی، خدمات مالی

۱. دانشجوی دکترای حسابداری، دانشگاه آزاد اسلامی واحد تهران جنوب، تهران، ایران.

۲. عضو هیات علمی دانشکده علوم مالی، دانشگاه خوارزمی، تهران، ایران. (نویسنده مسئول: rasoul.mostajeran@khu.ac.ir)

۱) مقدمه

مدل‌های یادگیری ماشین در حال تقویت کاربردهای هوش مصنوعی در بسیاری از حوزه‌ها، مانند خدمات مالی، مراقبت‌های بهداشتی و تولید هستند که عمدتاً به دلیل مزیت آن‌ها از نظر دقت پیش‌بینی نسبت به مدل‌های یادگیری آماری کلاسیک است. با این حال، اگرچه مدل‌های یادگیری ماشین می‌توانند به عملکرد بالایی در پیش‌بینی دست یابند، اما دارای طبیعت ذاتی غیرشفاف ("جعبه سیاه") هستند. این مسئله در صنایع تحت نظارت از جمله صنعت خدمات مالی مشکل‌ساز است، زیرا مقامات نظارتی که به دنبال پایش ریسک‌های ناشی از کاربرد هوش مصنوعی هستند، ممکن است نتوانند آن‌ها را تأیید کنند (بریک^۱ و همکاران، ۲۰۱۹).

ویژگی‌های یک مدل یادگیری ماشین مطلوب در صنعت خدمات مالی شامل چندین جنبه کلیدی است که به بهبود عملکرد، دقت و قابلیت اعتماد مدل‌ها کمک می‌کند. اولاً، دقت و صحت مدل از اهمیت بالایی برخوردار است؛ زیرا تصمیم‌گیری‌های مالی بر اساس پیش‌بینی‌های مدل باید دقیق و قابل اعتماد باشند تا از ریسک‌های مالی جلوگیری شود. ثانیاً، قابلیت تفسیر و توضیح‌پذیری مدل نیز حائز اهمیت است؛ به طوری که ذینفعان بتوانند دلایل تصمیمات مدل را درک کنند و در صورت نیاز، توجیهات لازم را ارائه دهند. همچنین، پایداری و مقاومت مدل در برابر تغییرات داده‌ها و شرایط بازار نیز از ویژگی‌های ضروری است، به طوری که مدل بتواند در طول زمان عملکرد خود را حفظ کند. انصاف و عدم تبعیض نیز باید در طراحی مدل‌ها مدنظر قرار گیرد تا از بروز تبعیض‌های ناعادلانه در تصمیم‌گیری‌ها جلوگیری شود. این ویژگی‌ها به طور کلی باعث می‌شوند که مدل‌های یادگیری ماشین بتوانند به طور مؤثری در صنعت خدمات مالی به کار گرفته شوند و تصمیم‌گیری‌های بهتری را تسهیل کنند (گیدیکی^۲ و همکاران، ۲۰۲۴).

ویژگی‌های یک مدل یادگیری ماشین مطلوب که تحت عنوان الزامات این مدل‌ها در مقررات جدید قرار گرفته‌اند را می‌توان در تهیه چارچوبی برای مدیریت ریسک کاربردهای هوش مصنوعی مورد استفاده قرار داد. به عبارتی اگر در کاربرد هوش مصنوعی این الزامات رعایت نشود، از اعتبار و قابلیت اطمینان کافی برخوردار نیست. بنابراین، باید روش‌های آماری برای این الزامات معرفی شوند تا بتوان آن‌ها را بر اساس این روش‌ها در کاربرد هوش مصنوعی اندازه‌گیری کرد و بر اساس آن یک چارچوب مدیریت ریسک یکپارچه تدوین کرد (گیدیکی و همکاران، ۲۰۲۴). در این پژوهش با تمرکز بر چهار ویژگی یک مدل یادگیری ماشین مطلوب یعنی پایداری، دقت، انصاف و توضیح‌پذیری، تلاش شده است تا برخی از روش‌های آماری اندازه‌گیری این ویژگی‌ها معرفی شوند تا بر اساس آن یک چارچوب مدیریت ریسک یکپارچه در زمینه کاربردهای هوش مصنوعی در صنعت خدمات مالی، معرفی شود.

اخیراً، چندین مقاله به بررسی کاربرد هوش مصنوعی برای اندازه‌گیری ریسک‌های خاص پرداخته‌اند. از جمله، نایم^۳ (۲۰۲۲) و سوندرا^۴ و همکاران (۲۰۲۳) که به کاربرد هوش مصنوعی در مدیریت ریسک مالی پرداخته‌اند؛ باکر^۵ و همکاران (۲۰۲۲)، بوسمان^۶ و همکاران (۲۰۲۰)، لیو^۷ و همکاران (۲۰۲۲)، موسکاتو^۸ و همکاران (۲۰۲۱) و سچان^۱ و

1. Bracke

2. Giudici

3. Naim

4. Sundra et al.

5. Bucker et al.

6. Bussmann et al.

7. Liu

8. Moscato et al.

همکاران (۲۰۲۰) که از هوش مصنوعی برای اندازه‌گیری ریسک اعتباری استفاده کرده‌اند؛ گیدیکی و ابو‌حشیش^۲ (۲۰۱۹) و گیدیکی و پولینزی^۳ (۲۰۲۱) که از هوش مصنوعی برای اندازه‌گیری ریسک‌های سرایت در معاملات الگوریتمی و صرافی‌های رمزارز استفاده کرده‌اند؛ گانش و کالبانا^۴ (۲۰۲۲) که روش‌های هوش مصنوعی را برای مدیریت ریسک زنجیره تأمین مرور کرده‌اند؛ فراست^۵ و همکاران (۲۰۱۹) و کوپیر^۶ و همکاران (۲۰۲۱) که روش‌های هوش مصنوعی را برای مدیریت ریسک‌های واسطه‌گری مالی انجام داده‌اند؛ آلداسورو^۷ و همکاران (۲۰۲۲)، گیدیکی و رافینتی^۸ (۲۰۲۱)، مک‌کال^۹ و همکاران (۲۰۱۷) و ملانکون^{۱۰} و همکاران (۲۰۲۱) که از هوش مصنوعی برای اندازه‌گیری ریسک‌های فناوری اطلاعات و ریسک‌های سایبری استفاده کرده‌اند؛ آپچک زای و خوان^{۱۱} (۲۰۲۲) و چن^{۱۲} و همکاران (۲۰۱۸) که از هوش مصنوعی برای شناسایی تقلب‌های مالی و پولشویی بهره برده‌اند.

با این حال، ادبیات قبلی بیشتر در حوزه ارائه روش‌های هوش مصنوعی برای اندازه‌گیری انواع مختلف ریسک‌ها و غالباً از منظر مالی توسعه یافته‌اند و کمتر به ریسک‌هایی که خود هوش مصنوعی ممکن است ایجاد کند، اشاره شده است. این در حالی است که خود این فناوری‌ها نیز می‌توانند عامل ایجاد ریسک‌هایی باشند و با توجه به استفاده گسترده کنونی از هوش مصنوعی، نیاز به توسعه روش‌های پیشرفته‌ای وجود دارد که بتوانند ریسک‌های آن را اندازه‌گیری کنند. از این منظر، این پژوهش که تلاش کرده است تا یک چارچوبی برای مدیریت ریسک‌های هوش مصنوعی در صنعت خدمات مالی ارائه دهد، یک پژوهش متمایز و منحصر به فرد است.

برای توسعه چنین چارچوبی، می‌توان به مقررات مالی که منجر به شناسایی ریسک‌هایی در صنعت خدمات مالی شده‌اند، اشاره کرد. زمانی که چارچوب سرمایه بازل II منتشر شد، ریسک‌های بازار، اعتباری و عملیاتی به عنوان متغیرهای کلیدی برای ارزیابی کفایت سرمایه یک مؤسسه مالی معرفی شدند و بعداً، با بازنگری بازل III (بانک تسویه‌حساب‌های بین‌المللی، ۲۰۱۱) نوبت به ریسک سیستمی رسید. در این میان، معیارهای آماری مانند ارزش در معرض ریسک^{۱۳} و افت مورد انتظار^{۱۴} (آرتزرنر^{۱۵} و همکاران، ۲۰۰۱) و بعداً، CoVaR (آدریان و برونرمیر^{۱۶}، ۲۰۱۶) توسط محققان پیشنهاد شده و سپس توسط بانک‌ها و نهادهای نظارتی برای پایش و اندازه‌گیری یکپارچه ریسک‌های مالی و پوشش آن‌ها توسط سرمایه داخلی بانک‌ها و مدیریت ریسک ادغام شده است.

به طور مشابه با آنچه پس از مقررات بازل اتفاق افتاد، در اینجا نیز می‌توان با استفاده از مقررات پیشنهادی هوش مصنوعی که به تازگی ارائه شده و چهار معیار پایداری، دقت، انصاف و توضیح‌پذیری را به عنوان ویژگی‌های مطلوب معرفی

¹. Sachan

². Giudici and Abu-Hashish

³. Giudici and Polinesi

⁴. Ganesh and Kalpana

⁵. Frostet al.

⁶. Kuiper

⁷. Aldasoro

⁸. Giudici and Raffinetti

⁹. McCall

¹⁰. Melancon

¹¹. Achakzai and Juan

¹². Chen

¹³. Value at Risk

¹⁴. Expected Shortfall

¹⁵. Artzner

¹⁶. Adrian & Brunnermeier

کرده‌اند، برای اندازه‌گیری ریسک‌های هوش مصنوعی در صنعت خدمات مالی اقدام کرد. این ویژگی‌ها نیاز به توسعه روش‌های آماری مناسب دارند تا بتوان مشابه با ریسک مالی یک مؤسسه مالی در مقررات بازل، به یک اندازه‌گیری یکپارچه از قابلیت اعتماد برای یک کاربرد خاص هوش مصنوعی پرداخت. پایداری و دقت، دو معیار فنی و داخلی و انصاف و توضیح‌پذیری، دو معیار اخلاقی و خارجی در مورد فرآیند هوش مصنوعی هستند و شامل ذینفعان یک سیستم هوش مصنوعی می‌شوند. توسعه چنین معیارهایی این امکان را فراهم می‌کند که نه تنها مشخص شود آیا یک کاربرد هوش مصنوعی قابل اعتماد است یا خیر، بلکه وضعیت قابلیت اعتماد را در طول زمان از طریق یک مدل مدیریت ریسک پایش می‌کند (گیدیکی و همکاران، ۲۰۲۴).

در این پژوهش، الزامات نظارتی جدیدی که برای هوش مصنوعی پیشنهاد شده مانند پایداری، دقت، انصاف، توضیح‌پذیری، در نظر گرفته شده و روش‌های آماری برای اندازه‌گیری آن‌ها پیشنهاد شده مه می‌توان از آن‌ها برای اندازه‌گیری، مدیریت و کاهش ریسک‌های ناشی از هوش مصنوعی استفاده کرد. امید است که این چارچوب پیشنهادی بتواند به طور مؤثر ریسک‌های هوش مصنوعی را اندازه‌گیری کند و به این ترتیب، هوش مصنوعی ایمن و قابل اعتماد را در حوزه مالی ترویج دهد. مهم‌ترین پیامد این پژوهش، ارائه یک مدل یکپارچه مدیریت ریسک برای کاربردهای هوش مصنوعی است.

(۲) مرور ادبیات

۲-۱) کاربردهای هوش مصنوعی در صنعت خدمات مالی

در سال‌های اخیر، پیشرفت‌های نرم‌افزاری و سخت‌افزاری باعث تسریع رشد فناوری‌های هوش مصنوعی در حوزه مالی شده است. این بخش به‌ویژه شاهد افزایش قابل توجهی در کاربردهای یادگیری ماشین بوده که به بهبود نتایج برای مصرف‌کنندگان و کسب‌وکارها منجر شده است. تا چند سال پیش، تنها صندوق‌های سرمایه‌گذاری از فناوری‌های هوش مصنوعی و یادگیری ماشین بهره‌مند بودند، اما اکنون این فناوری‌ها در بانک‌ها، فین‌تک‌ها، نهادهای قانونی و شرکت‌های بیمه نیز گسترش یافته‌اند. شرکت‌های فعال در صنعت خدمات مالی، از جمله بانک‌ها و بیمه‌ها، به سرعت الگوریتم‌های یادگیری ماشین را برای اتوماسیون فرآیندهای زمان‌بر و ارائه تجربه‌های شخصی‌سازی شده به مشتریان به کار می‌گیرند. این الگوریتم‌ها با استخراج بینش‌های معنادار از داده‌های خام، نتایج دقیقی ارائه می‌دهند و اطلاعات استخراج شده برای حل مسائل پیچیده در بانکداری و مالی استفاده می‌شود. همچنین، با گذشت زمان، این الگوریتم‌ها با یادگیری از داده‌ها و فرآیندهای موجود، بهبود یافته و بینش‌های جدیدتری از داده‌ها را کشف می‌کنند (پاتنایک^۱ و همکاران، ۲۰۲۴).

هوش مصنوعی به عنوان ابزار ارزشمندی در مبارزه با جرایم مالی، از جمله پول‌شویی و جرایم سایبری ظهور کرده است. مؤسسات مالی به منظور رعایت مقررات و جلوگیری از فعالیت‌های غیرقانونی، اقدامات ضد پول‌شویی را در اولویت قرار می‌دهند. با اتوماسیون تشخیص تراکنش‌های مشکوک، تکنیک‌های هوش مصنوعی، کارایی را افزایش داده و دخالت دستی را به حداقل می‌رسانند. علاوه بر این، این فناوری‌ها امکان شناسایی الگوهای رفتاری مشتریان که

¹. Pattnaik

نشان‌دهنده فعالیت‌های تقلبی هستند را فراهم می‌آورند و به این ترتیب اقدام‌های پیشگیرانه را تسهیل می‌کنند (فارمباخر^۱ و همکاران، ۲۰۲۲).

هوش مصنوعی همچنین کاربردهای گسترده‌ای در بخش بیمه دارد، به‌ویژه در زمینه‌های ذخیره‌سازی ادعاها و ارزیابی ریسک. از طریق تحلیل داده‌ها و مدل‌سازی پیش‌بینی، این فناوری به تسهیل قیمت‌گذاری دقیق و ارزیابی ریسک‌ها کمک می‌کند. علاوه بر این، چت‌بات‌ها و دستیاران مجازی مبتنی بر هوش مصنوعی، تجربه مشتری را با ارائه توصیه‌ها و کمک‌های متناسب بهبود می‌بخشند (بوهیا^۲ و همکاران، ۲۰۲۲). ملاحظه می‌شود که هوش مصنوعی کاربردهای متنوعی در صنعت خدمات مالی دارد که به بهبود کارایی و امنیت این بخش کمک می‌کند. در ادامه به برخی از این کاربردها اشاره شده است:

نظارت مالی: الگوریتم‌های یادگیری ماشین می‌توانند به تقویت امنیت شبکه‌ها کمک کنند. این الگوریتم‌ها با یادگیری تکنیک‌های جدید پول‌شویی، به نظارت دقیق‌تری منجر می‌شوند و در آینده نقش مهمی در امنیت سایبری ایفا خواهند کرد.

اتوماسیون فرایندها: استفاده از راه‌حل‌های یادگیری ماشین به مؤسسات مالی این امکان را می‌دهد که کارهای دستی و تکراری را خودکار کنند. این شامل چت‌بات‌ها، اتوماسیون کاغذی و گیمیفیکیشن فرایندهاست که به بهبود تجربه مشتری، کاهش هزینه‌ها و افزایش خدمات کمک می‌کند.

تراکنش‌های امن: الگوریتم‌های یادگیری ماشین در شناسایی تقلب‌های تراکنشی بسیار مؤثر هستند. این الگوریتم‌ها با تجزیه و تحلیل میلیون‌ها داده، رفتارهای مشکوک را شناسایی کرده و به جلوگیری از کلاهبرداری کمک می‌کنند. به عنوان مثال، سیستم‌های نظارتی بانک‌ها بر اساس تاریخچه تراکنش‌ها آموزش دیده‌اند تا تقلب را شناسایی کنند.

مدیریت ریسک: تکنیک‌های یادگیری ماشین به مؤسسات مالی کمک می‌کنند تا با تحلیل داده‌های گسترده، سطح ریسک را کاهش دهند. این الگوریتم‌ها می‌توانند اطلاعات شخصی بیشتری را نسبت به روش‌های سنتی تحلیل کنند و به تصمیم‌گیری‌های بهتر در اعطای وام کمک کنند.

مدیریت داده‌های مشتری: داده‌ها برای بانک‌ها و مؤسسات مالی حیاتی هستند و مدیریت مؤثر آن‌ها به رشد این کسب‌وکارها کمک می‌کند. یادگیری ماشین می‌تواند به پردازش و تحلیل حجم بالای داده‌ها کمک کند و بینش‌های ارزشمندی برای بهبود سودآوری ارائه دهد.

تصمیم‌گیری: بانک‌ها می‌توانند از الگوریتم‌های یادگیری ماشین برای تحلیل داده‌های ساختاریافته و بدون ساختار استفاده کنند. این الگوریتم‌ها به کشف روندهایی کمک می‌کنند که می‌تواند در تصمیم‌گیری‌های مالی مؤثر باشد.

بهبود سطح خدمات مشتری: با استفاده از چت‌بات‌های هوشمند، مشتریان می‌توانند به سؤالات خود پاسخ بگیرند و از خدمات شخصی‌سازی شده بهره‌مند شوند. همچنین، الگوریتم‌های یادگیری ماشین می‌توانند رفتار کاربران را تحلیل کرده و پیشنهادات سفارشی ارائه دهند.

¹. Farbmacher

². Bouhia

برنامه حفظ مشتری: شرکت‌های کارت‌های اعتباری می‌توانند با استفاده از یادگیری ماشین، مشتریان در معرض خطر را شناسایی کرده و اقدامات لازم برای حفظ آن‌ها را انجام دهند. این نرم‌افزارها با تحلیل داده‌های جمعیت‌شناختی و فعالیت‌های تراکنشی، پیشنهادات مناسبی برای حفظ مشتریان ارائه می‌دهند (ژو^۱ و همکاران، ۲۰۲۴).

در مجموع، هوش مصنوعی و یادگیری ماشین به مؤسسات مالی کمک می‌کنند تا کارایی، امنیت و تجربه مشتری را بهبود بخشند و در عین حال هزینه‌ها را کاهش دهند؛ بنابراین، هوش مصنوعی و یادگیری ماشین، صنعت خدمات مالی را متحول کرده‌اند و مدیریت ریسک را بهبود بخشیده‌اند (کائو^۲ و همکاران، ۲۰۲۲).

با این حال، خود این فناوری‌ها می‌تواند با ایجاد ریسک‌هایی همراه باشد؛ این ریسک‌ها شامل مسائل حریم خصوصی است، زیرا استفاده از داده‌های شخصی برای آموزش الگوریتم‌ها می‌تواند به نقض حریم خصوصی مشتریان منجر شود. همچنین، سوگیری الگوریتمی ممکن است به نتایج ناعادلانه‌ای منجر شود، به‌ویژه اگر داده‌های ورودی شامل سوگیری اجتماعی یا اقتصادی باشند. اعتماد به سیستم نیز یک چالش مهم است، زیرا مشتریان ممکن است به استفاده از سیستم‌های هوش مصنوعی اعتماد نکنند و فقدان شفافیت در نحوه کارکرد این الگوریتم‌ها می‌تواند نگرانی‌هایی ایجاد کند. علاوه بر این، خطرات امنیت سایبری وجود دارد، زیرا سیستم‌های هوش مصنوعی خود می‌توانند هدف حملات سایبری قرار گیرند. همچنین، اتوماسیون فرایندها ممکن است به کاهش مشاغل در بخش مالی منجر شود و این موضوع می‌تواند به بیکاری و نارضایتی اجتماعی منجر شود. پیچیدگی‌های قانونی و نظارتی نیز می‌تواند چالش‌هایی را به همراه داشته باشد، زیرا قوانین موجود ممکن است به‌روز نباشند و نتوانند به درستی به مسائل مرتبط با هوش مصنوعی پاسخ دهند. در نهایت، افزایش وابستگی به سیستم‌های هوش مصنوعی می‌تواند خطرات ناشی از نقص‌های فنی را افزایش دهد. برای بهره‌برداری مؤثر از هوش مصنوعی و یادگیری ماشین در صنعت مالی، ضروری است که مؤسسات به این ریسک‌ها توجه کنند و تدابیر لازم را برای مدیریت آن‌ها اتخاذ نمایند، از جمله توسعه یک چارچوب یکپارچه برای مدیریت ریسک‌های هوش مصنوعی که شاخص‌های کلیدی ریسک آن بر اساس روش‌های آماری اندازه‌گیری می‌شوند (گیدیکی و همکاران، ۲۰۲۴).

پیشرفت فناوری‌های جدید موجب رشد قابل توجهی در مباحث علمی پیرامون کاربردهای هوش مصنوعی و یادگیری ماشین در صنعت خدمات مالی شده است. تلاش‌های تحقیقاتی بر روی الگوریتم‌ها، مدل‌ها، ابزارها و چارچوب‌ها در زمینه‌های خاصی مانند خدمات مالی متمرکز شده است (گودل^۳ و همکاران، ۲۰۲۱). با این حال، بحث ریسک‌های هوش مصنوعی و اندازه‌گیری آن‌ها به منظور ایجاد یک چارچوب یکپارچه برای مدیریت ریسک هوش مصنوعی نادیده گرفته شده است.

۲-۲) توسعه چارچوب پیشنهادی برای مدیریت ریسک‌های هوش مصنوعی

مجموعه‌ای از شاخص‌های کلیدی برای مدیریت ریسک هوش مصنوعی که به مؤسسات مالی در اندازه‌گیری، مدیریت و کاهش ریسک‌ها در طول کل چرخه زندگی هوش مصنوعی، از طراحی تا نظارت مستمر کمک می‌کنند، عبارتند از پایداری، دقت، انصاف و توضیح‌پذیری.

¹. Zhu

². Cao

³. Goodell

الف) مفهوم پایداری^۱ و روش‌های آماری اندازه‌گیری آن

پایداری به تاب‌آوری خروجی‌های هوش مصنوعی در برابر رویدادهای غیرعادی و/یا حملات سایبری اشاره دارد. پیش‌بینی‌های یک مدل یادگیری ماشین، به‌ویژه زمانی که تعداد زیادی متغیر توضیحی در نظر گرفته می‌شود، ممکن است تحت تأثیر وجود نقاط داده «افراطی» قرار گیرند که ناشی از رویدادهای غیرعادی یا دستکاری داده‌های سایبری هستند. برای بهبود استحکام و امنیت سائیری برنامه‌های هوش مصنوعی (پایداری آن‌ها)، پیشنهاد می‌شود که روش‌های انتخاب متغیر موجود برای مدل‌های احتمالی، مانند رگرسیون خطی، به مدل‌های غیراحتمالی، مانند جنگل‌های تصادفی و مدل‌های شبکه عصبی، گسترش یابد. با انجام این کار، یک مدل کم‌حجم‌تر به دست می‌آید که در حالی که به‌طور معناداری دقت پیش‌بینی را از دست نمی‌دهد، ساده‌تر و بنابراین، در برابر تغییرات افراطی در داده‌ها پایدارتر خواهد بود. برای مدل‌های احتمالی، تفاوت در احتمال بین دو مدل تو در تو قابل محاسبه است و می‌توان یک آزمون آماری، مانند آزمون F برای پاسخ‌های پیوسته یا آزمون Chi-square برای پاسخ‌های باینری، را پیاده‌سازی کرد تا ارزیابی شود که آیا تفاوت بین این دو مدل، معنادار است یا خیر. به عنوان مثال، اگر مدل‌ها از منظر معکوس مقایسه شوند، به طوری که یک متغیر در هر بار حذف شود و از مدل کاملاً پارامتر شده (با بالاترین احتمال) شروع شود، این آزمون می‌تواند یک قاعده توقف ارائه دهد که بر اساس p-values آزمون است. این روند می‌تواند به مدلی منجر شود که در عین دقت، کم‌حجم‌تر و بنابراین پایدارتر از مدل کامل باشد.

برای مدل‌های غیر احتمالی، یک روند مرحله‌ای مشابه پیشنهاد می‌شود، اما متغیرها نه بر اساس احتمال آن‌ها که قابل محاسبه نیست، بلکه بر اساس یک معیار توضیح‌پذیری، مانند مقادیر شاپلی^۲، مرتب می‌شوند. سپس پیشنهاد می‌شود که یک قاعده توقف بر اساس مقایسه دقت پیش‌بینی، چه در مورد آزمون دیبولد و ماریانو^۳ (۱۹۹۵) (زمانی که متغیر پاسخ، پیوسته است) و چه در مورد آزمون دیلونگ^۴ و همکاران (۱۹۸۸) (زمانی که متغیر پاسخ، باینری است) به کار گرفته شود. قاعده توقف به این معنی است که متغیرها به همان اندازه که پیش‌بینی‌ها متفاوت هستند، اضافه می‌شوند، یعنی تا زمانی که تفاوت در دقت پیش‌بینی بین دو مدل متوالی بزرگ باشد و p-values کمتر از یک آستانه تعیین شده باشد. این معیارهای آمای، معیارهای انتخاب مدل برای مدل‌های یادگیری ماشین (غیر احتمالی) را فراهم می‌کند.

ب) مفهوم دقت^۵ و روش‌های آماری اندازه‌گیری آن

دقت به دقت پیش‌بینی خروجی‌های مدل اشاره دارد. اندازه‌گیری دقت پیش‌بینی بر اساس مقایسه بین شواهد پیش‌بینی‌شده و مشاهده‌شده انجام می‌شود. به طور معمول، مدل‌های ساده‌تر (مانند مدل‌های رگرسیونی) نسبت به مدل‌های پیچیده‌تر (مانند مدل‌های جنگل تصادفی) دقت کمتری دارند.

برای اندازه‌گیری دقت، باید تفاوت بین متغیرهای پاسخ پیوسته و متغیرهای پاسخ دسته‌ای مشخص شود. زمانی که متغیر پاسخ هدف، پیوسته است، می‌توان از معیارهایی مانند ریشه میانگین مربعات خطا (RMSE) برای پیش‌بینی‌ها استفاده کرد. برای استفاده از RMSE به عنوان یک شاخص کلیدی ریسک، پیشنهاد می‌شود که آن را با یک آزمون آماری، مانند آزمون دیبولد و ماریانو همراه کرد که می‌توان از آن برای تصمیم‌گیری در مورد اینکه آیا پیش‌بینی‌های دو مدل

^۱. Sustainability

^۲. Shapley values

^۳. Diebold and Mariano

^۴. DeLong

^۵. Accuracy

رقیب به طور معناداری متفاوت هستند یا خیر، استفاده کرد. هرچه RMSE کمتر باشد، مدل بهتر است و زمانی که p -values مربوطه کمتر از یک آستانه تعیین شده (مانند ۵٪) باشد، برابری پیش‌بینی‌ها رد می‌شود، یعنی مدل‌ها به طور معناداری متفاوت هستند.

زمانی که متغیر پاسخ هدف، دسته‌ای است، می‌توان از معیارهایی مبتنی بر مثبت‌های کاذب و منفی‌های کاذب، مانند مساحت زیر منحنی ROC (AUROC)، استفاده کرد. برای استفاده از AUROC به عنوان یک شاخص کلیدی ریسک، یک آزمون آماری، مانند آزمون دیلونگ و همکاران (۱۹۸۸)، پیشنهاد می‌شود که با آن بتوان تصمیم گرفت که آیا دسته‌بندی‌های دو مدل رقیب به طور معناداری متفاوت هستند یا خیر. منحنی ROC، برای مجموعه‌ای از نقاط قطع مشخص، نرخ مثبت واقعی را در برابر نرخ مثبت کاذب نمایش می‌دهد. هرچه AUROC بیشتر باشد، مدل بهتر است و برای یک مقدار p -values کوچک، برابری پیش‌بینی‌ها رد می‌شود.

ج) مفهوم انصاف^۱ و روش‌های آماری اندازه‌گیری آن

انصاف به عدم وجود سوگیری نسبت به گروه‌های جمعیتی ناشی از خروجی‌های هوش مصنوعی اشاره دارد. انصاف ویژگی‌ای است که به طور اساسی نیاز دارد که برنامه‌های هوش مصنوعی در میان گروه‌های مختلف جمعیتی سوگیری را نشان ندهند. بررسی انصاف باید در هر دو سطح داده (ورودی) و پیش‌بینی (خروجی) انجام شود. برای اندازه‌گیری انصاف، پیشنهاد می‌شود که ضریب جینی، که برای اندازه‌گیری تمرکز درآمد در یک جمعیت است، به اندازه‌گیری تمرکز متغیرهای توضیحی که ممکن است تحت تأثیر سوگیری قرار گیرند، گسترش یابد. این رویکرد امکان می‌دهد که نه تنها وجود سوگیری شناسایی شود، بلکه به طور دقیق‌تر درک شود که چگونه متغیرهای مختلف در تصمیم‌گیری‌های مدل تأثیر می‌گذارند و آیا این تأثیرات به طور عادلانه بین گروه‌های مختلف جمعیتی توزیع شده‌اند یا خیر. با استفاده از این معیار، می‌توان به بهبود انصاف در مدل‌های هوش مصنوعی کمک و اطمینان حاصل کرد که تصمیمات اتخاذ شده توسط این سیستم‌ها عادلانه و بدون سوگیری هستند.

این اندازه‌گیری می‌تواند بر اساس پارامترهای تخمینی (برای یک مدل قابل توضیح) یا بر اساس مقادیر شاپلی (برای یک مدل جعبه سیاه) باشد. برای یک متغیر توضیحی مشخص، نسبت‌های مشابه منجر به یک ضریب جینی نزدیک به ۰ می‌شود که نشان‌دهنده این است که تأثیر آن متغیر در بین گروه‌های جمعیتی مختلف عادلانه است. از سوی دیگر، یک ضریب جینی نزدیک به ۱ نشان می‌دهد که تأثیر یک متغیر به شدت به برخی گروه‌ها وابسته است، که نشان‌دهنده سوگیری و غیرعادلانه بودن آن است. در چارچوب مدیریت ریسک هوش مصنوعی پیشنهادی، ضریب جینی برای یک متغیر یک مقدار در بازه [۰, ۱] است که می‌تواند به عنوان یک شاخص کلیدی ریسک استفاده شود که درجه انصاف آن را نشان می‌دهد، به طوری که مقادیر بالاتر نشان‌دهنده انصاف بیشتر هستند.

شاخص کلیدی اندازه‌گیری ریسک پیشنهادی برای انصاف، میزان تمرکز قابلیت توضیح‌پذیری یک متغیر را در بین گروه‌های جمعیتی مختلف اندازه‌گیری می‌کند. برای ارزیابی اینکه آیا تمرکز به طور معناداری از توزیع یکنواخت متفاوت است، که نشان‌دهنده انصاف است، می‌توان شاخص کلیدی ریسک را با یک آزمون آماری غیرپارامتریک، مانند آزمون کولموگروف-اسمیرنوف، همراه کرد. یک p -values کوچک از آزمون نشان‌دهنده انحراف معنادار از توزیع یکنواخت است که نشان‌دهنده انصاف است.

¹. Fairness

د) مفهوم توضیح‌پذیری^۱ و روش‌های آماری اندازه‌گیری آن

توضیح‌پذیری به قابلیت فهم و نظارت بر خروجی‌های مدل توسط انسان‌ها، به‌ویژه در زمینه علل محرکه آن، اشاره دارد. توضیح‌پذیری تأثیر مثبتی بر تمامی ذینفعان مرتبط با یک سیستم هوش مصنوعی دارد. اگر سیستم قابل تفسیر باشد، یک دانشمند داده می‌تواند مدیریت بهتری بر تکامل‌های فنی مدل داشته باشد. حساب‌رسان و نهادهای نظارتی نیز می‌توانند مدل‌های هوش مصنوعی را بهتر تأیید کنند. همچنین، مشتریان نهایی دوست دارند از دلایل پیش‌بینی‌هایی که شامل آن‌ها می‌شود (مانند درخواست‌های وام اعتباری) و نحوه پردازش داده‌هایشان مطلع شوند. در حالی که برخی از مدل‌های هوش مصنوعی از نظر پارامترهایشان (مانند مدل‌های رگرسیونی)، به‌طور ذاتی قابل توضیح هستند، برخی دیگر به عنوان جعبه‌های سیاه شناخته می‌شوند (مانند شبکه‌های عصبی و جنگل‌های تصادفی).

برای مدل‌های جعبه سیاه، پیشنهاد می‌شود که از مقادیر شاپلی استفاده شود، که توسط شاپلی (۱۹۵۳) معرفی شده و توسط بریک و همکاران (۲۰۱۹) و لوندبرگ و لی^۲ (۲۰۱۷) به زمینه هوش مصنوعی قابل توضیح تطبیق داده شده است. مقادیر شاپلی در زمینه نظریه بازی‌ها معرفی شده‌اند. اگر فرض شود برای هر مشاهده‌ای که باید پیش‌بینی شود، یک بازی وجود دارد؛ برای هر بازی، بازیکنان پیش‌بینی‌کننده‌های مدل هستند و کل سود برابر با مقدار پیش‌بینی شده است که به عنوان مجموع سهم هر پیش‌بینی‌کننده به دست می‌آید.

پس از محاسبه مقادیر شاپلی، برای هر مشاهده، سهم جهانی هر متغیر توضیحی در پیش‌بینی‌ها با جمع یا میانگین‌گیری آن‌ها بر روی تمام مشاهدات به دست می‌آید. برای مدل مدیریت ریسک پیشنهادی در این پژوهش، به عنوان یک شاخص کلیدی ریسک، درصد مقادیر شاپلی متغیرهای موجود در مدل یادگیری ماشین انتخاب شده پیشنهاد می‌شود. اگر مدل به طور کامل پارامتر شده باشد، شاخص کلیدی ریسک برابر با ۱۰۰٪ است؛ اما در شرایط واقعی یک مدل کم‌حجم‌تر و قابل توضیح‌تر، شاخص کلیدی ریسک مقداری بین ۰٪ و ۱۰۰٪ خواهد داشت.

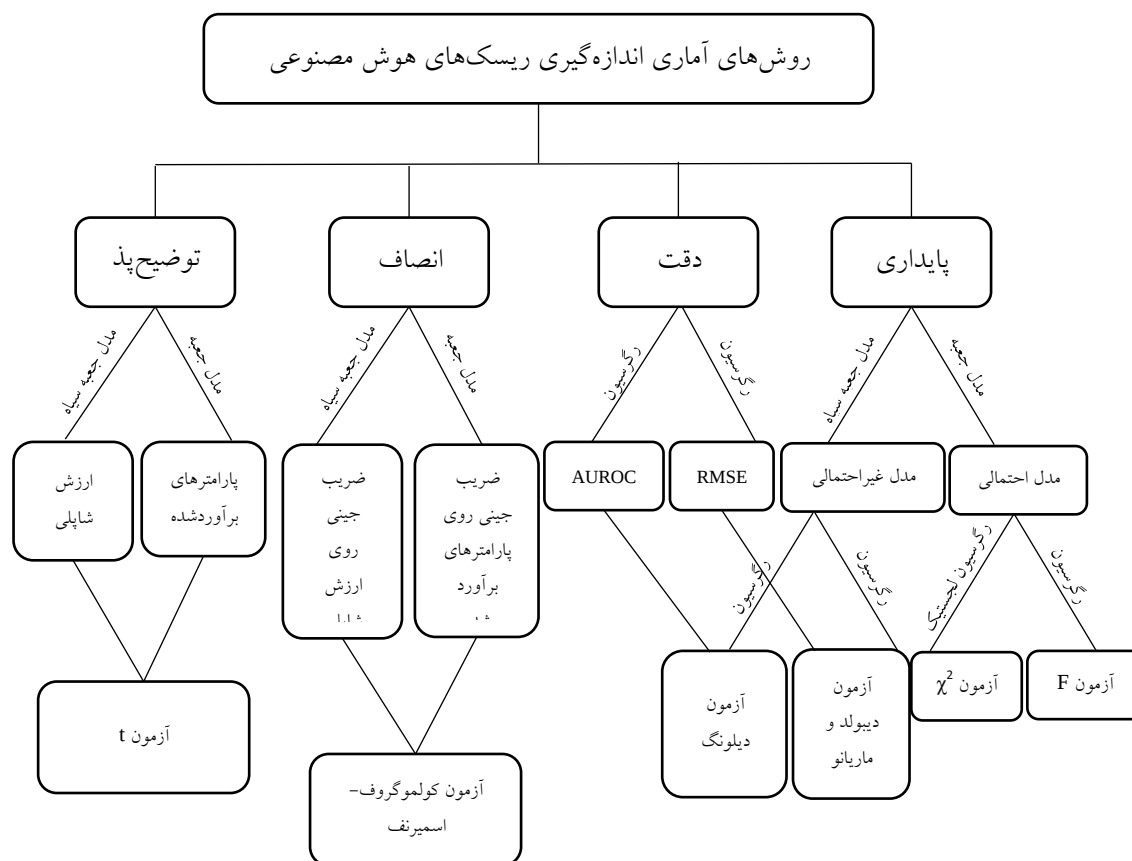
یادآوری می‌شود که برای یک مدل شفاف (جعبه سفید)، برای ارزیابی اینکه آیا توضیح‌پذیری یک متغیر از نظر آماری معنادار است، می‌توان از آزمون t برای پارامترهای تخمینی مربوطه استفاده کرد. یک روند مشابه می‌تواند برای یک مدل جعبه سیاه اعمال شود، به طوری که متغیرهای توضیحی با مقادیر شاپلی مربوطه جایگزین شوند. در هر دو حالت، یک p -values کوچک از آزمون نشان‌دهنده توضیح‌پذیری معنادار خواهد بود.

با توجه به مطالب فوق، چارچوب پیشنهادی اندازه‌گیری ریسک‌های هوش مصنوعی در صنعت خدمات مالی به صورت شکل (۱) ارائه می‌شود.

با پیاده‌سازی اقدامات پیشنهادی در چارچوب مدیریت ریسک، ناظران می‌توانند به طور سیستماتیک پایداری، دقت، انصاف و توضیح‌پذیری مدل‌های مختلف یادگیری ماشین را ارزیابی و بهبود بخشند. این رویکرد جامع نه تنها در انتخاب و اعتبارسنجی مدل کمک می‌کند، بلکه اطمینان حاصل می‌کند که سیستم‌های هوش مصنوعی مسئولانه و با استانداردهای اخلاقی همسو هستند. هر بُعد به درک جامع‌تری از عملکرد مدل و تأثیر آن بر ذینفعان کمک می‌کند و در نهایت اعتماد بیشتری در مورد کاربردهای هوش مصنوعی ایجاد می‌کند.

^۱. Explainability

^۲. Lundberg and Lee



شکل ۱- چارچوب پیشنهادی اندازه‌گیری ریسک‌های هوش مصنوعی

۳-۲) پیشینه تجربی پژوهش

پژوهش‌های بسیار محدودی در زمینه ریسک‌های هوش مصنوعی در صنعت خدمات مالی انجام شده است و بیشتر ادبیات قبلی بر روی کاربردهای هوش مصنوعی و یا اندازه‌گیری ریسک‌های خاص با هوش مصنوعی تمرکز داشته‌اند. در جدول (۱) برخی از این پژوهش‌ها و یافته‌های آن‌ها مشاهده می‌شود. برای یافتن این پژوهش‌ها به جستجوی کلیدواژه‌های مرتبط همچون، ریسک‌های هوش مصنوعی، صنعت خدمات مالی، کاربردهای هوش مصنوعی و اندازه‌گیری ریسک‌های مالی، در پایگاه منابع اینترنتی پرداخته شده است.

جدول ۱- مرور بر سوابق موضوع پژوهش

ردیف	پژوهشگر/سال	عنوان	نتایج
۱	ترابی و رجبی فرجاد (۱۴۰۳)	مدل هوش مصنوعی پیشگوی بازرسی جهت کشف فساد و ریسک	استفاده از هوش مصنوعی و یادگیری ماشین در فرآیندهای بازرسی و کشف فساد، نه تنها دقت بالایی در شناسایی فساد دارد، بلکه می‌تواند به عنوان یک ابزار پیشگوی ریسک، سازمان‌ها را در مدیریت بهتر منابع و کاهش ریسک‌ها یاری دهد.
۲	وقفی و دارابی (۱۳۹۸)	اعتبارسنجی الگوریتم‌های هوش	الگوریتم درخت تصمیم با استفاده از داده‌های مالی و اقتصادی کارایی بالاتری نسبت به روش طبقه‌بندی نیز و

جدول ۱- مرور بر سوابق موضوع پژوهش

ردیف	پژوهشگر/سال	عنوان	نتایج
		مصنوعی در پیش‌بینی درماندگی مالی	ماشین بردار پشتیبان در جهت پیش‌بینی درماندگی مالی دارد.
۳	نعیم ^۱ و همکاران (۲۰۲۵)	نقش هوش مصنوعی در مدیریت ریسک بخش بانکی	هوش مصنوعی به طور مستقیم بر مدیریت ریسک تأثیر می‌گذارد و رابطه مثبتی بین هوش مصنوعی و مدیریت ریسک وجود دارد.
۴	گیدی و همکاران (۲۰۲۴)	اندازه‌گیری ریسک‌های هوش مصنوعی	چارچوب پیشنهادی اندازه‌گیری ریسک‌های هوش مصنوعی می‌تواند به طور مؤثری ریسک‌های هوش مصنوعی را اندازه‌گیری کند و به این ترتیب، هوش مصنوعی ایمن و قابل اعتمادی را در حوزه خدمات مالی ترویج دهد.
۵	باکر و همکاران (۲۰۲۲)	شفافیت، قابلیت حسابرسی و توضیح‌پذیری مدل‌های یادگیری ماشین در امتیازدهی اعتباری	این مقاله ابعاد مختلفی را که باید برای قابل فهم کردن مدل‌های امتیازدهی اعتباری در نظر گرفته شوند، بررسی می‌کند و چارچوبی برای شفافیت، قابلیت حسابرسی و توضیح‌پذیری مدل‌های "جعبه سیاه" یادگیری ماشین ارائه می‌دهد.
۶	دیباک و التاجمان ^۲ (۲۰۲۱)	حفظ حریم خصوصی در قراردادهای هوشمند با استفاده از بلاک چین و هوش مصنوعی برای اندازه‌گیری ریسک‌های سایبری	این مقاله چارچوبی به نام حفظ حریم خصوصی در قراردادهای هوشمند با استفاده از بلاک چین و هوش مصنوعی را پیشنهاد می‌کند که تعاملات انسانی، فعالیت‌های سیستم، هشدارهای خدمات، ریسک‌های امنیتی و ادعاهای تقلبی را ساده می‌کند. برای تحلیل داده‌های تراکنش و اشتراک‌گذاری، از روش تقویت گرادیان افراطی استفاده می‌شود. این روش خدمات تراکنش را تحلیل می‌کند تا بار شبکه را بهینه کند و نشان دهد که آیا نرخ انتقال در ارتباط با شبکه‌های غیرقابل اعتماد تغییر می‌کند یا خیر.
۷	کوپر و	بررسی هوش	برای موارد استفاده اعتبار مصرف‌کننده، ریسک اعتباری و

^۱. Naeem^۲. Deebak and AL-Turjman

جدول ۱- مرور بر سوابق موضوع پژوهش

ردیف	پژوهشگر/سال	عنوان	نتایج
	همکاران (۲۰۲۱)	مصنوعی قابل توضیح در بخش مالی	مبارزه با پول‌شویی، تفاوت‌هایی بین نهادهای نظارتی و بانک‌ها در مورد دامنه مطلوب توضیح‌پذیری سیستم‌های هوش مصنوعی وجود دارد.
۸	بوسمان و همکاران (۲۰۲۰)	کاربرد هوش مصنوعی در مدیریت ریسک اعتباری	هم وام‌گیرندگان پرخطر و هم وام‌گیرندگان کم‌خطر می‌توانند بر اساس مجموعه‌ای از ویژگی‌های مالی مشابه گروه‌بندی شوند. این ویژگی‌ها می‌توانند برای توضیح نمره اعتباری آن‌ها و در نتیجه پیش‌بینی رفتار آینده‌شان به کار گرفته شوند.

ملاحظه می‌شود که پژوهش‌ها در زمینه اندازه‌گیری و مدیریت ریسک‌های هوش مصنوعی در صنعت خدمات مالی، بسیار پراکنده و محدود هستند و پتانسیل بررسی‌های پژوهشی بیشتری در این راستا به منظور پر کردن شکاف مطالعاتی و غنی کردن ادبیات موجود و همچنین ارائه کاربردهای عملی وجود دارد.

(۳) بحث و نتیجه‌گیری

در دنیای امروز، هوش مصنوعی به عنوان یک ابزار کلیدی در صنعت خدمات مالی شناخته می‌شود که می‌تواند به بهبود کارایی، دقت و سرعت در فرآیندهای مختلف کمک کند. با این حال، استفاده از هوش مصنوعی در این صنعت با چالش‌ها و ریسک‌های متعددی همراه است که نیاز به یک رویکرد جامع و سیستماتیک برای مدیریت این ریسک‌ها را ضروری می‌سازد. در این پژوهش، چارچوبی به منظور اندازه‌گیری و مدیریت ریسک‌های هوش مصنوعی در صنعت خدمات مالی معرفی شده است که می‌تواند به نهادهای مالی کمک کند تا از مزایای هوش مصنوعی بهره‌برداری کنند، در حالی که ریسک‌های مرتبط با آن را به حداقل برسانند.

مدیریت ریسک‌های هوش مصنوعی در صنعت خدمات مالی از اهمیت ویژه‌ای برخوردار است. ریسک‌های مربوط به هوش مصنوعی می‌توانند شامل ریسک‌های عملیاتی، ریسک‌های قانونی، ریسک‌های اخلاقی و ریسک‌های امنیتی باشند. با توجه به اینکه تصمیمات مالی بر اساس داده‌ها و الگوریتم‌های هوش مصنوعی اتخاذ می‌شوند، وجود پایداری، دقت، انصاف و توضیح‌پذیری در این فرآیندها ضروری است که قانون جدید نظارت بر کاربردهای هوش مصنوعی به عنوان الزامات و ویژگی‌های مطلوب یک مدل یادگیری ماشین از آن‌ها یاد شده است. در این راستا، این الزامات می‌تواند به عنوان شاخص‌های کلیدی ریسک هوش مصنوعی در نظر گرفته شوند که عدم وجود آن‌ها کاربردهای هوش مصنوعی را فاقد اعتبار کرده و قابلیت اطمینان آن را کاهش می‌دهد. بنابراین در این پژوهش پیشنهاد شده است که روش‌های آماری برای اندازه‌گیری این معیارها در یک کاربرد هوش مصنوعی در صنعت خدمات مالی (به عنوان مثال امتیازدهی اعتباری) باید تعیین شود تا این معیارها با استفاده از این روش‌های آماری سنجیده شوند و در نهایت بر اساس این معیارها و کارایی آن‌ها یک چارچوب مدیریت ریسک یکپارچه ایجاد شود. چارچوب پیشنهادی به نهادهای مالی کمک می‌کند تا با استفاده از اصول کلیدی پایداری، دقت، انصاف و توضیح‌پذیری، میزان ریسک‌های هوش مصنوعی را شناسایی و مدیریت کنند.

- چهار شاخص کلیدی که به شناسایی و مدیریت ریسک‌ها کمک می‌کند به شرح زیر هستند:
- **پایداری:** اطمینان از اینکه سیستم‌های هوش مصنوعی به طور مداوم و در طول زمان عملکرد مناسبی دارند و تحت تأثیر تغییرات محیطی قرار نمی‌گیرند.
 - **دقت:** تضمین اینکه مدل‌های هوش مصنوعی به درستی و با دقت بالا پیش‌بینی‌ها و تصمیمات لازم را انجام می‌دهند.
 - **انصاف:** جلوگیری از تبعیض و نابرابری در تصمیمات اتخاذ شده توسط الگوریتم‌های هوش مصنوعی.
 - **توضیح‌پذیری:** فراهم کردن امکان فهم و تفسیر تصمیمات گرفته شده توسط مدل‌های هوش مصنوعی برای کاربران و نهادهای نظارتی.
- این چهار اصل به عنوان مبنایی برای توسعه شاخص‌های کلیدی ریسک هوش مصنوعی عمل می‌کنند که می‌توانند به نهادهای مالی در ارزیابی و مدیریت ریسک‌های مربوط به هوش مصنوعی کمک کنند.
- این پژوهش توضیح داده است که با برخی از روش‌های آماری می‌توان این شاخص‌های کلیدی را اندازه‌گیری کرد و با پیاده‌سازی چارچوب مدیریت ریسک‌های هوش مصنوعی در چهار مورد استفاده کلیدی در صنعت خدمات مالی—از جمله امتیازدهی اعتباری، مبارزه با پول‌شویی، نظارت بر سیستم‌های IT و تشخیص ناهنجاری—می‌توان از اصول ایمنی برای بهبود فرآیندهای تصمیم‌گیری و کاهش ریسک‌ها استفاده کرد. این کاربردها نه تنها به افزایش کارایی و دقت کمک می‌کنند، بلکه به نهادهای مالی امکان می‌دهند تا با اطمینان بیشتری از فناوری‌های هوش مصنوعی بهره‌برداری کنند.
- این پژوهش، یک چارچوب مدیریت ریسک هوش مصنوعی یکپارچه پیشنهاد داده است که می‌تواند با رعایت مقررات جدید هوش مصنوعی، ریسک‌های کاربردهای هوش مصنوعی را ارزیابی کند. این چارچوب از چهار اصل کلی که باید با روش‌های آماری اندازه‌گیری شوند، استفاده می‌کند. در این پژوهش، برای هر اصل، مجموعه‌ای از روش‌های آماری یکپارچه معرفی شدند که می‌توانند برای اندازه‌گیری ریسک‌های هوش مصنوعی و پیاده‌سازی یک سیستم مدیریت ریسک هوش مصنوعی مؤثر برای هر کاربرد هوش مصنوعی استفاده شوند.
- چارچوب مدیریت ریسک هوش مصنوعی پیشنهادی می‌تواند از توسعه‌دهندگان و کاربران هوش مصنوعی، همچنین ناظران و مقامات نظارتی در توسعه و نظارت بر مدل‌های یادگیری ماشین حمایت کند که با رعایت الزامات قانونی، می‌توانند ایمنی و قابلیت اعتماد کاربردهای هوش مصنوعی را ارزیابی کرده و ریسک‌های آن‌ها را اندازه‌گیری کنند.
- در نهایت، توسعه چارچوبی برای مدیریت ریسک‌های هوش مصنوعی در صنعت خدمات مالی نه تنها به نهادهای مالی کمک می‌کند تا از مزایای هوش مصنوعی بهره‌برداری کنند، بلکه به ایجاد یک محیط امن و قابل اعتماد برای استفاده از این فناوری نیز کمک می‌کند. با توجه به سرعت پیشرفت فناوری‌های هوش مصنوعی و تأثیر آن‌ها بر صنعت مالی، این چارچوب می‌تواند به عنوان یک راهنمای عملی برای نهادهای مالی در جهت مدیریت ریسک‌ها و اطمینان از انطباق با الزامات قانونی و اخلاقی عمل کند. به این ترتیب، انتظار می‌رود که این پژوهش به گسترش دانش و درک بهتری از چالش‌ها و فرصت‌های هوش مصنوعی در صنعت خدمات مالی کمک کند و مسیرهای جدیدی برای تحقیقات آینده فراهم آورد.

۴) پیشنهادها

با توجه به گسترش روزافزون کاربرد هوش مصنوعی در صنعت خدمات مالی، مدیران و برنامه‌ریزان در این صنعت باید توجه داشته باشند که بدون مدیریت ریسک‌های هوش مصنوعی، ممکن است اختلالاتی در فرایند انجام خدمات مالی پدید آید که اعتماد کاربران این صنعت را سلب کند؛ بنابراین، باید زمینه‌های استفاده بدون نقض از این فناوری از جمله مدیریت ریسک‌های آن، در ابتدای امر مورد توجه قرار گیرد و سپس کاربردهای آن توسعه یابد. با توجه به اصولی که در این پژوهش به عنوان ریسک‌های هوش مصنوعی به آن‌ها اشاره شده است، می‌توان به متولیان قانون‌گذار در این حوزه، پیشنهاد داد که به منظور ایجاد یک چارچوب قانونی و نظارتی مؤثر بر کاربردهای هوش مصنوعی، بر میزان پایداری، دقت، انصاف و توضیح‌پذیری این فناوری در خدمات مالی تمرکز داشته باشند و یک رویکرد یکپارچه برای پایش و مدیریت ریسک‌های هوش مصنوعی به موسسات مالی معرفی کنند تا اطمینان حاصل شود که هوش مصنوعی به شیوه‌ای امن و مؤثر در صنعت خدمات مالی به کار گرفته می‌شود و در عین حال اعتماد کاربران و ذینفعان حفظ می‌شود.

با توجه به اینکه پژوهش حاضر یک پژوهش ترویجی و بر اساس یافته‌های قبلی است، بدون اینکه شواهد تجربی جدیدی ارائه دهد، می‌تواند زمینه پژوهش‌های تجربی را فراهم کند که در قالب پیشنهادی زیر به پژوهشگران ارائه شده است.

پژوهشگران می‌توانند از منظر ریاضی و آماری پژوهش‌هایی را در صنعت خدمات مالی در محیط ایران، انجام دهند تا مفاهیم مختلف ارائه شده در شکل (۱) را به هم متصل کرده و راه‌های جایگزینی برای اندازه‌گیری اصول تعیین شده و آزمایش معنی‌داری آماری آن‌ها پیشنهاد کنند.

پژوهش‌های دیگری می‌تواند چارچوب پیشنهادی را به زمینه‌های دیگر غیر از صنعت خدمات مالی اعمال کند، جایی که یک هوش مصنوعی ایمن می‌تواند ارزش داده‌ها را افزایش دهد و در عین حال ریسک‌های ناشی از آن را کنترل کند.

منابع

ترابی، محمدامین؛ رجبی فرجاد، حاجیه. (۱۴۰۳). مدل هوش مصنوعی پیشگوی بازرسی جهت کشف فساد و ریسک. نظارت و بازرسی، ۱۸(۶۸)، ۴۵-۸۴.

وقفی، سید حسام؛ دارابی، رویا. (۱۳۹۸). اعتبارسنجی الگوریتم‌های هوش مصنوعی در پیش‌بینی درماندگی مالی در بخش صنعت و معدن با تأکید بر نقش متغیرهای کلان اقتصادی، مالی، مدیریتی و ریسک. پژوهشنامه بازرگانی، ۲۳(۹۱)، ۲۴۳-۲۱۳.

Achakzai, M. A. K., & Juan, P. (2022). Using machine learning meta-classifiers to detect financial frauds. *Finance Research Letters*, 48, Article 102915.

Adrian, T., & Brunnermeier, M. (2016). CoVaR. *American Economic Review*, 166(7), 1705-1741.

Aldasoro, I., Gambacorta, L., Giudici, P., & Leach, T. (2022). The drivers of cyber risk. *Journal of financial stability*, 60, Article 100989.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (2001). Coherent measures of risk. *Mathematical Finance*, 9(3), 203-228.

Bouhria, M., Rajaobelina, L., PromTep, S., Arcand, M., & Ricard, L. (2022). Drivers of privacy concerns when interacting with a chatbot in a customer service encounter. *International Journal of Bank Marketing*, 40, 1159-1181.

- Bracke, P., Datta, A., Jung, C., & Sen, S. (2019). *Machine learning explainability in finance: an application to default risk analysis: Staff working paper n. 816*, London: Bank of England.
- Bücker, M., Szepannek, G., Gosiewska, A., & Biecek, P. (2022). Transparency, auditability, and explainability of machine learning models in credit scoring. *Journal of the Operational Research Society*, 73(1), 70–90.
- Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). Explainable machine learning in credit risk management. *Computational Economics*, 57, 2013–2016.
- Cao, T., Cook, W. D., & Kristal, M. M. (2022). Has the technological investment been worth it? Assessing the aggregate efficiency of non-homogeneous bank holding companies in the digital age. *Technological Forecasting and Social Change*, 178, 121576.
- Chen, Z., Van Khoa, L. D., & Teoh, E. N. (2018). Machine learning techniques for antimoney laundering (AML) solutions in suspicious transaction detection: a review. *Knowledge in Information Systems*, 57, 245–285.
- Deebak, B. D., AL-Turjman, F. (2021). Privacy-preserving in smart contracts using blockchain and artificial intelligence for cyber risk measurements. *Journal of Information Security and Applications*, 58, 102749.
- Farbmacher, H., Löw, L., & Spindler, M. (2022). An explainable attention network for fraud detection in claims management. *Journal of Economic Behavior & Organization*, 228, 244–258.
- Frost, J., Gambacorta, L., Huang, Y., & Zbindnen, P. (2019). BigTech and the changing structure of financial intermediation. *Economic Policy*, 34(100), 761–799.
- Ganesh, A. D., & Kalpana, P. (2022). Future of artificial intelligence and its influence on supply chain risk management – A systematic review. *Computers & Industrial Engineering*, 169.
- Giudici, P., & Abu-Hashish, I. (2019). What determines bitcoin exchange prices? A network VAR approach. *Finance Research Letters*, 28, 309–318.
- Giudici, P., & Polinesi, G. (2021). Crypto price discovery through correlation networks. *Annals of Operations Research*, 299(1–2), 443–457.
- Giudici, P., & Raffinetti, E. (2021). Explainable AI in cyber risk management. *Quality and Reliability Engineering International*, 38(3), 1318–1326.
- Giudici, P., Centurelli, M., Turchetta, S. (2024). Artificial Intelligence risk measurement. *Expert Systems with Applications*, 235, 121220.
- Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance*, 32, 100577.
- Kuiper, O., Berg, M. Van. den., Burgt, J. Van. der., & Leijnen, S. (2021). Exploring explainable AI in the financial sector: Perspectives of banks and supervisory authorities. In *Artificial intelligence and machine learning*. Springer.
- Liu, W., Fan, H., & Xi, M. (2022). Credit scoring based on tree-enhanced gradient boosting decision trees. *Expert Systems with Applications*, 189, Article 116034.
- Lundberg, S. M., & Lee, S. (2017). A unified approach to interpreting model predictions. In *Proceedings of neural information processing systems 2017* (pp. 4768–4777).
- McCall, J. Ainslie., Shakya, S., & Owusu, G. (2017). Predicting service levels using neural networks. In *International conference on innovative techniques and applications of artificial intelligence* (pp. 411–416). Springer.
- Melancon, G., Grangier, P., Prescott-Gagnon, E., Sabourin, E., & Rousseau, L. M. (2021). A machine learning-based system for predicting service-level failures in supply chains. *INFORMS Journal on Applied Analytics*, 51(3), 200–212, 2021.

- Moscato, V., Picariello, A., & Sperlí, G. (2021). A benchmark of machine learning approaches for credit score prediction. *Expert Systems with Applications*, 165, Article 113986.
- Naeem, M., Siraj, M., Ali, S., Rehman, A., Farooq, S. (2025). The Role of Artificial Intelligence in Risk Management: Practices of the Banking Sector. *Generative AI for Web Engineering Models*.
- Naim, A. (2022). Role of artificial intelligence in business risk management. *American Journal of Business Management, Economics and Banking*, 1, 55–66.
- Pattnaik, D., Ray, S., Raman, R. (2024). Applications of artificial intelligence and machine learning in the financial services industry: A bibliometric review. *Heliyon*, 10(1), e23492.
- Sachan, S., Yang, Y., Xu, E., & Li, D. (2020). An explainable AI decision-supportsystem to automate loan underwriting. *Expert Systems with Applications*, 144, Article 113100.
- Sundra, B. M., Sathiyamurthi, K., & Subramanian, G. (2023). Critical evaluation of applying machine learning approaches for better handling bank risk management in the post-modern era. *Scandinavian Journal of Information Systems*, 35(1), 1228–1231.
- Zhu, H., Vigren, O., Söderberg, I-L. Implementing artificial intelligence empowered financial advisory services: A literature review and critical research agenda. *Journal of Business Research*, 174, 114494.