فصلنامـه اقتصاد مالی
دورهٔ ۱۹ / شمارهٔ ۳ (پیاپی ۷۲)/ پائیز ۱٤۰٤
صفحه ٤٥۹ تا ٤۹۹

کشف صورت های مالی متقلبانه براساس تکنیک یادگیری عمیق حافظه طولانی کوتاه مدت (LSTM)

جواد شرف خانی[۱]
محمدعلی بیداری[۲]✉

تاریخ دریافت: ۱٤۰٤/۰٤/۲۹ تاریخ پذیرش: ۱٤۰٤/۰٦/۲۹

چکیده

هدف پژوهش حاضر کشف صورت های مالی متقلبانه براساس تکنیک یادگیری عمیق حافظه طولانی کوتاه مدت در شرکت های پذیرفته شده در بورس اوراق بهادار تهران و مقایسه نتایج حاصل از آن با سایر الگوریتم های یادگیری عمیق می باشد. به منظور نیل به این هدف ۱۸۰۰ سال-شرکت (۱۵۰ شرکت برای ۱۲ سال) مشاهده جمع آوری شده از گزارش های مالی سالیانه شرکت های پذیرفته شده در بورس اوراق بهادار تهران در طی دوره زمانی ۱۳۹۱ تا ۱۴۰۲ مورد آزمون قرار گرفته اند. در پژوهش حاضر از چهار الگوریتم یادگیری عمیق (شامل الگوریتم های حافظه طولانی کوتاه مدت (LSTM)، ماشین بردار پشتیبان(SVM)، شبکه عصبی پیچشی(CNN) و شبکه عصبی بازگشتی (RNN)) و همچنین جهت انتخاب متغیرهای نهایی پژوهش از روش آزمون مقایسه میانگین دو نمونه جهت ایجاد مدل استفاده شده است. نتایج حاصل از الگوریتم های یادگیری عمیق نشان می دهد که صحت کلی الگوریتم های LSTM، SVM، CNN و RNN به ترتیب 98.6٪، 88.4٪، 81.3٪ و 94.4٪ می باشد که نشان دهنده این است الگوریتم LSTM بهترین عملکرد و الگوریتم CNN بدترین عملکرد را در کشف شرکت های دارای صورت های مالی متقلبانه دارد. به عبارت دیگر، نتایج نشان دهنده کارآ بودن الگوریتم حافظه طولانی کوتاه مدت (LSTM) نسبت به سایر الگوریتم های یادگیری عمیق می باشد. بنابراین در شرکت های پذیرفته شده در بورس اوراق بهادار تهران الگوریتم LSTM کاراترین مدل را برای کشف صورت های مالی متقلبانه فراهم می کند. یافته های پژوهش حاضر می تواند برای سهامداران، اعتباردهندگان، تحلیل گران و سایر مشارکت کنندگان در بازار سرمایه اطلاعات سودمندی را در بهبود پیش بینی صورت های مالی متقلبانه و کاهش خطا در اتخاذ تصمیمات مبتنی بر اطلاعات صورت های مالی، ارزیابی بهتر عملکرد شرکت بر مبنای اطلاعات، اتخاذ استراتژی های معاملاتی بهینه و شناسایی فرصت های مناسب خرید و فروش سهام بر مبنای اطلاعات صورت های مالی فراهم کند.

واژه‌های کلیدی: صورت های مالی متقلبانه، کشف تقلب، یادگیری عمیق، الگوریتم حافظه طولانی کوتاه مدت
طبقه بندی JEL: M41,M42 ,G32

۱ گروه اقتصاد ، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران javad.sharafkhani@iau.ac.ir
۲ گروه اقتصاد، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران.(نویسنده مسئول) bidari@ut.ac.ir

نوع مقاله: علمی پژوهشی
Eci@iauctb.ac.ir

۱- مقدمه

تقلب در گزارشگری مالی در سال های اخیر رشد قابل توجهی داشته است، در این راستا صورت های مالی هرچه بیشتر مورد تحریف قرارگرفته است. در مواقعی که صورت های مالی حاوی تحریفی بااهمیت باشد به گونه ای که عناصر تشکیل دهنده آن صورت مالی بیانگر واقعیت نباشد، گفته می شود که تقلب صورت گرفته است (اسپاتیس[1]، 2002). لذا توجه به اهمیت کشف تقلب در صورت های مالی در جهت حمایت از منافع سرمایه گذاران، امری بسیارمهم است. متأسفانه تقلب از هر نوع آن و با هر قصدی تأثیر نامطلوبی بر شرکت خواهد داشت. از این رو شناخت و پیشگیری از آن امری ضروری برای شرکت ها محسوب می شود. اعمال متقلبانه به زیان های مالی بزرگی برای اقتصادها و واحدهای تجاری در سطح جهان منجر شده است. بنابراین در زمان کنونی، کشف و مبارزه با تقلب جهت جبران زیان های وارده بسیار بااهمیت است. صورت های مالی متقلبانه به عنوان حذفیات عمده یا ارائه نادرست در صورت های مالی که به دلیل عدم گزارش عمدی داده های مالی مطابق با استانداردهای حسابداری پذیرفته شده است، مشخص می شود. صورت های مالی متقلبانه می تواند تأثیرات قابل توجهی بر ذینفعان شرکت های متقلب و شرکت های غیر متقلب در صورت عدم شناسایی و پیشگیری به موقع ایجاد کند (ال-بانانی و همکاران[2]، ۲۰۲۰). متأسفانه تشخیص صورت های مالی متقلبانه آسان نیست. علاوه بر این، حتی در صورت کشف، آسیب قابل توجهی به طور کلی قبلاً رخ داده است. در نتیجه، سیاست گذاران، حسابرسان و سرمایه‌گذاران از تکنیک‌های کارآمدتر و مؤثرتری که می‌توانند صورت های مالی متقلبانه را شناسایی کنند، سود زیادی خواهند برد. انجمن بازرسان خبره تقلب[3] بیان می‌کند که تقلب در صورت‌های مالی عبارت است از ارائه نادرست عمدی وضعیت مالی یک شرکت از طریق تحریف یا حذف عمدی مبالغ یا افشا در صورت‌های مالی برای گمراه کردن استفاده‌کنندگان از صورت‌های مالی. براساس گزارش مرکز کیفیت حسابرسی۴، افراد یا شرکت‌ها به دلایل مختلفی از جمله مزایای پولی، نیاز به تحقق اهداف مالی کوتاه‌مدت یا سرپوش گذاشتن بر اخبار ناگوار درگیر دستکاری صورت‌های مالی هستند. استفاده کنندگان خارجی و داخلی صورت‌های مالی دائماً صورت‌های مالی را زیر سؤال می‌برند و نهادهای نظارتی نمی‌توانند با اطمینان بگویند که صورت‌های مالی معتبر هستند و مطابق با دستورات قانونی و اخلاقی رویه‌های حسابداران و حسابرسان تهیه شده‌اند (کولیکووا و ساتداروا[5]، ۲۰۱۶؛ دیبیاک و همکاران[6]، ۲۰۲۲). در نتیجه، کشف تقلب یا فریب به منظور اطمینان از صحت صورت های مالی مهم است. در این زمینه، مطالعه حاضر از اهمیت عملی برای مشاغل و حسابرسان برخوردار است، زیرا بازار جهانی شاهد افزایش تقلب در حسابداری مالی است که میلیاردها دلار در سال برای کسب و کارها هزینه دارد. آشفتگی مالی تأثیر قابل توجهی بر کسب و کارها و اعتباردهندگان یک کشور و در نتیجه بر اقتصاد آن دارد (ال-بانانی و همکاران، ۲۰۲۰؛

[1] *Spathis*
[2] El-Bannany et al.
[3] The Association of Certified Fraud Examiners (ACFE)
[4] The Center for Audit Quality (CAQ)
[5] Kulikova and Satdarova
[6] Deebak et al.

اسرییدهاران و همکاران[1]، ۲۰۲۰؛ کومار و تریپاتی[2]، ۲۰۲۰). در نتیجه، کشف و پیش‌بینی تقلب حسابداری مالی به موضوعی نوظهور برای مطالعات دانشگاهی و متخصصان صنعت تبدیل می‌شود.

در محیط تجاری مدرن امروزی با پیچیده تر شدن سیستم ها و فعالیت ها و ازدیاد اطلاعات فریبکارانه، فناوری های مورد استفاده برای جلوگیری و کشف تقلب نیز به روز شده است؛ بنابراین توسعه و رشد روش های مربوط به داده های متنوع از جمله داده کاوی برای تشخیص تقلب های مالی در اولویت می باشد. در داخل کشور به موضوع کشف تقلب صورت های مالی توجه زیادی نشده است. مسئله گزارشگری مالی متقلبانه در ایران از اهمیتی ویژه برخوردار است. افزایش تعداد شرکت های پذیرفته شده در بورس که به منظور جذب منابع مالی به انتشار اوراق بهادار اقدام می کنند، تلاش به منظور کاهش مالیات بر درآمد و ... از جمله دلیل اهمیت این موضوع است. بنابراین در حوزه کشف تقلب هرچقدر حسابرسان و ذینفعان شرکت به تکنیک ها و رویکردهای کارآمدتری از جمله تکنیک های یادگیری عمیق و رویکردهای کارآمدی چون تکنیک یادگیری عمیق حافظه طولانی کوتاه مدت (LSTM) دسترسی داشته باشند قاعدتاً، توانایی و کارایی آنها در حوزه تائیدپذیری اطلاعات صورت های مالی را می تواند بهبود ببخشد.

کشف تقلب در صورت های مالی به چند دلیل دشوار است؛ نخست، اعضای گروه مدیریت که عمدتاً مرتکبین اصلی تقلب هستند، به طور معمول تلاش زیادی برای پنهان ساختن تقلب در صورت های مالی انجام می دهند؛ دوم، به این دلیل که اعضای گروه مدیریت از اعتماد قابل توجهی برخوردار هستند، به راحتی کنترل های داخلی کلیدی را نقض می کنند و در نتیجه تقلب در صورت های مالی آسان تر و کشف آن سخت تر می شود؛ سوم، مرتکبین تقلب در صورت های مالی اغلب از تبانی و سند سازی برای ارنکاب و پنهان کردن تقلب استفاده می کنند و بیشتر رویه های حسابرسی مستقل برای کشف چنین طرح هایی از تقلب طراحی نشده اند؛ چهارم، داده های صورت های مالی تا حد زیادی خلاصه و تجمیع شده اند که پنهان کردن کردن تقلب را ساده تر و کشف آن را با استفاده از روش های مدل سازی آماری و تحلیلی سخت تر می کند (وایتینگ و همکاران[3]، ۲۰۱۲). دو جریان متفاوت در متون پژوهشی، پیشینه و مبنایی برای این پژوهش فراهم می کند؛ یکی پژوهش های انجام شده در رابطه با گزارشگری مالی متقلبانه و دیگری متون مربوط به تکنیک های یادگیری عمیق است. این دو جریان قلمروهای پژوهشی گسترده ای هستند که پژوهش های متقابل بسیار اندکی در داخل کشور، در این زمینه انجام شده است. پژوهش های صورت گرفته نیز با توجه به محدودیت در زمینه دسترسی به داده های تقلب با یکدیگر متفاوت هستند؛ از این رو، این پژوهش در تلاش است تا با استفاده از تکنیک یادگیری عمیق حافظه طولانی کوتاه مدت (LSTM) و مقایسه نتایج آن با سه تکنیک یادگیری ماشین (ماشین بردار پشتیبان، شبکه عصبی پیچشی و شبکه عصبی بازگشتی) و با بکارگیری جامع ترین متغیرهای اثرگذار بر تقلب در صورت های مالی (با بررسی جامع

[1] Sreedharan et al.
[2] Kumar et al.
[3] Whiting et al

پژوهش داخلی و خارجی انجام شده در حوزه تقلب) سعی در ارائه مدلی برای کشف صورت های مالی متقلبانه با دقت و صحت بالا می باشد، که این موضوع در نوع خود منحصربفرد می باشد.

با توجه به مطالب بیان شده، پژوهش حاضر به دنبال پاسخی به این سؤال است که چگونه تکنیک یادگیری عمیق حافظه طولانی کوتاه مدت (LSTM) قدرت کشف صورت های مالی متقلبانه را در شرکت های پذیرفته شده در بورس اوراق بهادار تهران را بهبود می بخشد؟ به عبارت دیگر، صحت و دقت تکنیک یادگیری عمیق حافظه طولانی کوتاه مدت (LSTM) در کشف صورت های مالی متقلبانه در شرکت های پذیرفته شده در بورس اوراق بهادار تهران چه میزان می باشد؟

۲. مبانی نظری و پیشینه پژوهش

۲-۱. مبانی نظری پژوهش

در سطح جهانی، هزینه صورت های مالی متقلبانه طی سال‌ها به افزایش خود ادامه داده است. مستقیم ترین قربانیان سرمایه گذاران و سرمایه دارانی هستند که به بهانه های واهی به شرکت ها وجوه می دهند. هزینه های بیشتری برای جامعه در سطح کلان وجود دارد، مانند از دست دادن اعتماد به سیستم های مالی، حتی اگر تأثیر اقتصادی در سطح شرکت می تواند بسیار متغیر باشد. از آنجایی که صورت های مالی متقلبانه آسیب مالی قابل توجهی به سرمایه گذاران، سهامداران و جامعه وارد می کند، تعداد زیادی از مطالعات انجام شده است. اکثر آثار موجود در ادبیات صورت های مالی متقلبانه از تحلیل رگرسیون سنتی(به عنوان مثال، اندرو و رابین[1]، ۲۰۲۲؛ پآولو سرگئو گومس ماسدا و ویئیرا[2]، ۲۰۲۲؛ وادوا و کومار[3]، ۲۰۲۰؛ آمر و جربوئی[4]، ۲۰۱۳؛ آلسینقلاوی و آلماری[5]، ۲۰۲۱) استفاده می کنند.

یادگیری ماشین که زیرمجموعه هوش مصنوعی است، به علت خاصیت اکتشافی خود، بدون هیچ فرض اولیه ای شروع به الگوسازی رفتار داده ها می کند و به مرور زمان و با جلو رفتن الگوریتم، الگو پررنگ تر خواهد شد؛ ساختار غیرخطی و مقاوم این روش، توانایی شبیه سازی رفتار محیط واقعی را فراهم کرده است (میتچل[6]، ۱۹۹۷). در یادگیری ماشینی از طریق به کارگیری یک سری از الگوریتم های خاص سعی می شود الگوهای پنهان بین داده ها کشف شود؛ در صورتی که تعداد متغیرهای مستقل و وابسته زیاد باشد و بین آنها ارتباط خطی برقرار نباشد، این روش در زمره بهترین گزینه ها برای پیش بینی است. یادگیری عمیق، زیرشاخه ای از یادگیری ماشینی و از حوزه های نوین این بخش است. یادگیری عمیق معرف سلسله ای از الگوریتم هاست که در آن از چندین لایه پردازش اطلاعات به ویژه اطلاعات غیرخطی استفاده می شود تا بهترین ویژگی های مناسب از ورودی خام ا ستخراج شود(دنگ و یو، ۲۰۱۴). از میان تکنیک های مختلف یادگیری عمیق که در علوم مختلف کاربردهای

[1] Andrew and Robin
[2] Paulo Sérgio Gomes Macedo and Vieira
[3] Wadhwa and Kumar
[4] Ama and Jarboui
[5] Alsinglaw I and Almari
[6] Mitchell

فراوانی دارد، پژوهشگران حوزه مالی برای پیش بینی تقلب، اغلب از الگوریتم های به خصوصی نظیر شبکه عصبی بازگشتی (RNN)، حافظه طولانی کوتاه مدت (LSTM) و شبکه عصبی پیچشی استفاده کرده اند (راسام[1]، ۲۰۲۰).

شبکه‌های عصبی مصنوعی به دلیل ماهیت همه‌کاره آن به طور گسترده برای حل بسیاری از مشکلات استفاده شده است. یک رویکرد ترکیبی، یعنی ترکیبی از متغیرهای تحلیل بنیادی و تکنیکی شاخص‌های بازار سهام برای پیش بینی قیمت سهام با هوش مصنوعی در آینده برای بهبود روش‌های موجود ارائه شد و حرکت شاخص قیمت سهام با استفاده از دو مدل مبتنی بر شبکه عصبی مصنوعی و ماشین بردار پشتیبان مورد بحث قرار گرفت. پس از مقایسه عملکرد هر دو مدل، نتیجه این شد که میانگین عملکرد مدل ANN به طور قابل‌توجهی بهتر از مدل SVMاست و اینکه شبکه عصبی مصنوعی تناسب بهتری با داده‌ها نسبت به روش‌های مرسوم فراهم می‌کند. این شبکه‌های عصبی به گونه‌ای توسعه یافته‌اند که می‌توانند الگوهایی را از داده‌های نویزدار ا استخراج کنند. ANN ابتدا یک سیستم را با استفاده از نمونه بزرگی از داده‌ها به نام فاز آموزشی آموزش می‌دهد، سپس شبکه را با داده‌هایی آشنا می‌کند که در مرحله آموزش گنجانده نشده است، این مرحله به عنوان مرحله اعتبارسنجی یا پیش‌بینی شناخته می‌شود. تنها انگیزه این روش پیش‌بینی نتایج جدید است. این ایده یادگیری از آموزش و سپس پیش‌بینی نتایج در ANN از مغز انسان می‌آید که می‌تواند یاد بگیرد و پاسخ دهد. بنابراین ANN در بسیاری از کاربردها مورد استفاده قرار گرفته است و در اجرای توابع پیچیده در زمینه‌های مختلف موفقیت‌آمیز ثابت شده است (جی و همکاران[2]، ۲۰۲۰).

LSTM، نوعی از شبکه‌های عصبی است که برای پردازش داده‌های سری زمانی طراحی شده است. این شبکه با استفاده از یک واحد حافظه خاص، می‌تواند اطلاعات را به مدت طولانی‌تری نگه‌داری کند تا در زمان‌های بعدی به آن‌ها دسترسی داشته باشد. در LSTM، هر نورون دارای وضعیت داخلی و حالت سلول است. وضعیت داخلی نورون برای حفظ و نگه‌داری اطلاعات قبلی از حافظه و حالت سلول برای حفظ و نگه‌داری اطلاعات جدید استفاده می‌شود. این دو وضعیت در هر مرحله از زمان، با استفاده از دروازه‌هایی مانند دروازه فراموشی، دروازه ورودی و دروازه خروجی کنترل می‌شوند(مِهتاب و همکاران[3]، ۲۰۲۰).

در حالت کلی، به دلیل پیچیدگی مسئله تقلب و حجم بالای اطلاعات مورد پردازش، اغلب استفاده از یک سیستم ساده برای پیش بینی نتایج خوبی به همراه ندارد. به همین دلیل محققان با ارائه ی مدلی ترکیبی سعی در ارائه ی سیستمی با پیچیدگی کمتر و کارایی و دقت بیشتر کرده اند. امروزه از الگوهای مختلفی مانند: تکنیک های آماری (تحلیل تشخیصی، لوجیت و آنالیز فاکتوری) و تکنیک های هوش مصنوعی (شبکه های عصبی، درخت تصمیم گیری، استدلال مبتنی بر موضوع، الگوریتم ژنتیک، مجموعه های سخت، ماشین بردار پشتیبان و منطق فازی) و یا ترکیبی از این دو تکنیک برای پیش بینی تقلب استفاده می شود. در اکثر مدل های پیش بینی کننده، سیستم فقط با استفاده از اطلاعات یک شاخص به پیش بینی می پردازد، اما در مدل پیشنهادی

[1] Rasam
[2] Jay et al.
[3] Mehtab et al.

در این پژوهش یک الگوریتم مبتنی بر سری زمانی شبکه عصبی بازگشتی با حافظه‌ی طولانی کوتاه مدت بهره گرفته می شود و جهت سنجش کارایی و دقت، نتایج آن با چندین الگوریتم یادگیری عمیق/پیش بینی دیگر (ماشین بردار پشتیبان، شبکه عصبی پیچشی و شبکه عصبی بازگشتی) مقایسه می گردد.

با ظهور یادگیری عمیق، مدل های جدید برای تحلیل و پیش بینی سریهای زمانی توسعه یافته اند. شبکه های عصبی مکرر(روملهارت و همکاران[1]، ۱۹۸۶) بیشتر، کارآمدترین روش پیش بینی سری زمانی اند(لیو و همکاران[2]، ۲۰۲۰). درواقع، RNN، شبکه مصنوعی است که در آن گره ها به صورت حلقه وصل می شوند و حالت داخلی شبکه می تواند رفتار زمان بندی پویا را به نمایش بگذارد. مهم ترین ضعف شبکه های مبتنی بر RNN در هنگام یادگیری وابستگی های طولانی مدت است. برای غلبه بر این مشکل، LSTM و واحدهای مکرر گیت دار[3] ارائه شدند که عملیات خطی ساده را روی اطلاعات نورون انجام می دهند و اطلاعات خارجی از لحظه فعلی را ازطریق مکانیزم گیت اضافه می کنند(چونگ و همکاران[4]، ۲۰۱۴). با وجود مزیت های شبکه های LSTM ، هنوز عملکرد آنها بر داده های پیش بینی سری زمانی، رضایت بخش نیست. معماری های کم عمق ویژگی های داده های سری زمانی را به طور کارآمد نشان نمی دهند؛ به خصوص زمانی که داده های سری زمانی با فواصل طولانی مدت و بسیار غیرخطی پردازش می شوند (سغیر و کتب[5]، ۲۰۱۹). همچنین با پیچیده تر شدن معماری شبکه های عمیق، یک سؤال پیش می آید که چطور یک شبکه را تنظیم کنیم؛ البته می توان تعداد محدودی از ابرپارامترها را با آزمایش بهینه کارد ؛ اما شبکه های عمیق دارای توپولوژی پیچیده و صدها ابرپارامترند. اغلب موفقیت در حل مسئله به انتخاب معماری مناسب برای آن مسئله بستگی دارد(میکولینن و همکاران[6]، ۲۰۱۹).

۲-۲. پیشینه تجربی پژوهش

ابوت و همکاران[7] (۲۰۰۰) استقلال حسابرسی و فعالیت در کاهش احتمال تقلب را بررسی و اندازه گیری نمودند.با استفاده از تحلیل رگرسیون لوجستیک آنها دریافتند که شرکتهای دارنده کمیته حسابرسی که ترکیبی از هیات مدیره غیرموظف و دارای جلساتی حداقل دوبار در سال هستند؛ تمایل کمتری به تصمیم گزارشگری متقلبانه یا گمراه کننده دارند. استاتیوس و همکاران (۲۰۰۷) با استفاده از نسبتهای مالی به عنوان متغیرهای ورودی و با بکارگیری روشهای داده کاوی نحوه کشف تقلب صورتهای مالی را بررسی کردند که مدل درخت تصمیم و شبکه عصبی و شبکه باور بیزین به ترتیب از نرخ صحت پیش بینی ۹۶ درصد و ۱۰۰درصد و ۹۵ درصد حکایت دارند.این نتایج بیانگر توانایی کشف تقلب از طریق داده های صورتهای مالی می باشد. لو و وانگ[8] (۲۰۰۹) به ارزیابی احتمال

[1] Rumelhart et al.
[2] Liu et al.
[3] Gated Recurrent Unit
[4] Chung et al.
[5] Sagheer and Kotb
[6] Miikkulainen et al
[7] Abbott et al
[8] Lu and Wang

گزارشگری مالی متقلبانه از طریق بررسی عوامل ریسک مثلث تقلب پرداختند و دریافتند که گزارشگری مالی متقلبانه همبستگی مثبتی با «فشارهای وارده بر شرکت، وجود درصد بالاتری از معاملات پیچیده در شرکت، زیر سؤال بودن صلاحیت و دستکاری مدیریت، عدم وجود رابطه ای مناسب و خوب بین مدیریت و حسابرس» دارد.
احمد و همکاران[1] (۲۰۰۹) به بررسی رابطه میان گزارش های مالی متقلبانه و ویژگی های شرکت(اندازه، نوع مالکیت و کیفیت حسابرسی) پرداختند. آن ها در مطالعه خود از روش رگرسیون حداقل مربعات و نظریه های هزینه سیاسی استفاده کرده اند. نتایج آنان نشان می دهد اندازه شرکت و کیفیت حسابرسی رابطه منفی و معناداری با تقلب در گزارشگری مالی دارد. پرولز و لوگی[2] (۲۰۱۱) به بررسی مشخصه های شرکت های متقلب پرداختند و این موضوع که آیا انجام مدیریت سود توسط شرکت بر احتمال وقوع تقلب در صورت های مالی آن تأثیر دارد یا خیر را مورد بررسی قرار دادند. نتایج آنها نشان داد که احتمال تقلب در شرکت هایی که در دوره های پیشین اقدام به مدیریت سود کرده اند، بیش تر است.

کسکی و هانلون[3] (۲۰۱۳) در پژوهشی رابطه بین گزارشگری مالی متقلبانه و سیاست تقسیم سود شرکت ها را مطالعه کردند. نتایج نشان داد که شرکت های متهم به تقلب، نسبت به شرکت های دیگر، سود نقدی کمتری پرداخت می کنند و سیاست تقسیم سود یکنواختی ندارند. آلبرچت و همکاران[4] (۲۰۱۵) درمقاله ای اثرقدرت بر گزارشگری مالی متقلبانه را بررسی کردند. آنان برای نمایش چگونگی استفاده ازقدرت درجلب مشارکت افراد دیگر در فرایند گزارشگری مالی از مدل رده بندی قدرت فرنچ راوان استفاده کردند. نتایج تحقیق آنان نشان می دهد که فرد الف با استفاده از تطمیع؛ تهدید؛ قدرت قانونی؛ مهارت حرفه ای وتوجیه فرد ب را به مشارکت درتقلب فرا میخواند. ونگ و همکاران[5] (۲۰۱۷) در پژوهشی با عنوان «توانایی مدیریتی، ارتباطات سیاسی و گزارشگری مالی متقلبانه در چین» براساس نمونه متشکل از شرکت های پذیرفته شده در بورس اوراق بهادار چین در طی دوره زمانی ۲۰۱۲-۲۰۰۷ به این نتیجه رسیدند که افزایش توانایی مدیریتی منجر به کاهش تقلب در گزارشگری مالی می گردد. نتایج همچنین نشان می دهد که ارتباطات سیاسی شرکت می تواند تأثیر توانایی مدیریتی بر گزارشگری مالی متقلبانه را تضعیف یا محدود کند.

ژانگ و چن[6] (۲۰۱۷) در پژوهشی با عنوان " سن هیئت مدیره و کلاهبرداری مالی شرکت ها: چشم انداز تعاملی" فرصتهای تقلب را با محرکهای موقعیتی و خصوصیات شخصی مدیران ارزیابی کرده وبه این نتیجه رسیده اند که هرچه میانگین سن هیئت مدیره افزایش یابد احتمال تقلب بیشتر است ولی در شرکت های بزرگتر این موضوع اهمیت کمتری دارد.و مدیرعامل قوی در شرکت های بزرگتر میتواند اثر سن هیئت مدیره را تضعیف کند.
تانگ و کریم[7] (۲۰۱۸) در پژوهشی با عنوان " کشف تقلب مالی و تجزیه و تحلیل داده های بزرگ – پیامدهای

[1] Ahmed et al
[2] Prowls and Logi
[3] Kesky and Hanlon
[4] Albrecht et al.
[5] Wong et al.
[6] Zhang and Chen
[7] Tang and Karim

مورد استفاده حسابرسان از نشست های ایده پردازی" به این نتیجه رسیدند که روش های حسابرسی موجود با هدف کشف تقلب نیاز به بهبود و اصلاح دارند چرا که توجه به خرد جمعی باعث ایجاد نگرانی و گمراهی خواهد شد بنابراین باید از داده های بزرگتر در هر مرحله از جمع آوری داده ها و شناسایی شاخص های تقلب و مستندات بیشتر و جلسات متعدد استفاده شود. رهروی دستجردی و فروغی (۲۰۱۸) پژوهشی با عنوان " کشف ریسک تقلب مدیران با استفاده از تجزیه و تحلیل متون : شواهدی از ایران" و با هدف مقایسه دقیق دو روش داده کاوی در تشخیص و کشف این خطر پرداختند به این منظور دو روش بهینه سازی محدب (CVX) و حداقل انقباض مطلق و انتخاب اپراتور (LASSO) روش رگرسیون بررسی کرده و مدلی را ارائه دادند که با دقت بالا میتواند شاخص ریسک تقلب مدیر را در شرکتها تشخیص دهد.

چن لیو و همکاران[1] (۲۰۱۸) در پژوهشی با عنوان "تشخیص تقلب برای صورتهای مالی گروه های تجاری"رویکردی را برای تشخیص کلاهبرداری در صورتهای مالی گروه های تجاری برای کاهش ضرر ریسک سرمایه گذاری و افزایش مزایای سرمایه گذاری با استفاده از طراحی فرایندی برای تشخیص تقلب و توسعه الگوریتم های کشف تقلب عنوان کردند. سادگالی و همکاران[2] (۲۰۱۹) در پژوهشی با عنوان "عملکرد الگوریتم های یادگیری ماشینی در تشخیص تقلب مالی " یک روش اثربخش و کارآمد در کشف تکنیکهای مختلف تقلب و همچنین الگوریتم های پیشگیری تقلب را با خوشه بندی و رگرسیون پیشنهاد کردند. وان هویلد و همکاران[3] (۲۰۱۹) در پژوهشی با عنوان "کشف تقلب ارزش افزوده مالیاتی با استفاده الگوریتم های ناهنجاری تشخیصی"نشان دادند که با روشهای پیشنهادی موارد تقلب قابل تشخیص می باشد (۱)تهیه فرم مالیات با لیست های مشتری و ارزش پیش بینی این ترکیب (۲) طراحی الگوریتم های سریع برای مقابله با ارقام بالای مالیات (۳) ارزیابی عملکرد کشف تقلب.

علی و همکاران[4] (۲۰۲۳) در پژوهشی از تکنیک‌های مختلف یادگیری ماشین در محیط پایتون برای پیش‌بینی صورت های مالی متقلبانه استفاده کردند و یافته‌های تجربی آنها نشان می‌دهد که الگوریتم XGBoost از دیگر الگوریتم‌های این مطالعه، یعنی رگرسیون لجستیک (LR)، درخت تصمیم (DT)، ماشین بردار پشتیبان (SVM)، جنگل تصادفی (RF) و آدابوست بهتر عمل کرده است. و سپس الگوریتم XGBoost را برای بدست آوردن بهترین نتیجه با دقت نهایی ۹۶.۰۵ درصد در تشخیص صورت های مالی متقلبانه را بهینه کردند. زهیر و همکاران[5] (۲۰۲۳) در پژوهشی با استفاده از داده های شاخص ترکیبی شانگهای و روش های CNN، RNN، LSTM، CNN-RNN و CNN-LSTM نشان دادند که CNN بدترین عملکرد را دارد، LSTM بهتر از CNN-LSTM، CNN-RNN بهتر از CNN-LSTM و CNN-RNN بهتر از LSTM عمل می‌کند، و مدل RNN تک لایه پیشنهادی بر همه مدل‌های

[1] Chen Liu et al.
[2] Sadgali et al
[3] Van Hoveld et al
[4] Ali et al
[5] Zaheer et al.

اشاره شده برتری دارد. نتایج تجربی اثربخشی مدل پیشنهادی را تایید می‌کند که به سرمایه‌گذاران در افزایش سود خود با تصمیم‌گیری خوب کمک می‌کند.

فرقاندوست حقیقی و برواری(۱۳۸۸) در تحقیقی با عنوان بررسی کاربرد روشهای تحلیلی در ارزیابی ریسک تقلب صورتهای مالی(تقلب مدیریت) ملاک طبقه بندی ارتکاب به تقلب یا عدم تقلب را وجود اظهارنظر مردود یا عدم اظهارنظر دانسته و با استفاده ازتعداد ۱۲نمونه نهایی و درنظرگرفتن ۷ نسبت مالی به عنوان متغیرهای مستقل؛ دو متغیر نسبت بدهی ها به داراییها و نسبت رشد فروش معنادار تشخیص داده شده و درصد صحت پیش بینی مدل ۹۰ درصد ذکر شده است. صفرزاده (۱۳۸۹) در پژوهشی با عنوان «توانایی نسبتهای مالی در کشف تقلب در گزارشگری مالی: تحلیل لاجیت» به تحلیل اطلاعات ۱۷۸ شرکت در طی دوره زمانی ۱۳۸۳ تا ۱۳۸۶ پرداختند و پس از تحلیل، ده نسبت مالی به عنوان پیش‌بینی کننده‌های بالقوه گزارشگری مالی متقلبانه معرفی شد. نتایج تحقیق حکایت از عملکرد مناسب الگو در طبقه‌بندی شرکتهای نمونه داشت به‌گونه‌ای که درصد صحت طبقه‌بندی الگو از ۹۸.۸۲ درصد تجاوز نمود. هم‌چنین نتایج نشان داد که الگوی تحقیق، توانایی کشف تقلب در گزارشگری مالی را دارد و الگوی پیشنهادی می‌تواند به گروه‌های مختلف استفاده کننده همچون حسابرسان، مقامات مجاز مالیاتی، سیستم بانکی و ... در تمایز شرکت های متقلب از غیرمتقلب کمک کنند. پورحیدری وبذرافشان (۱۳۹۰) اقدام به رتبه بندی بسترهای خطر تقلب باتوجه به میزان اهمیت آن ها در کشف تقلب مدیریت پرداختند. نتایج تحقیق آنان نشان می دهد؛ مهمترین بستر خطرتقلب؛"وابسته بودن بخش عمده ای از حقوق ومزایای مدیران به نتایج عملیات ؛وضعیت مالی یا جریان وجوه نقد" است.

مرادی و همکاران(۱۳۹۳) در پژوهشی با عنوان «شناسایی عوامل خطر مؤثر بر احتمال وقوع تقلب در گزارشگری مالی از دید حسابرسان و بررسی تأثیر آن ها بر عملکرد مالی شرکت» به این نتیجه رسیدند که بین ویژگی های مدیریت، تبعیت مدیریت از کنترل های داخلی و استانداردهای لازم الاجرا، عوامل خطر مرتبط با شرایط بازار و صنعت، ویژگی های عملیاتی، نقدینگی و ثبات مالی با احتمال وقوع تقلب رابطه ای معناداری وجود دارد. هم چنین نتایج حاکی از وجود رابطه ای معنادار بین عملکرد شرکت(نرخ بازده دارایی ها، جریان های نقدی عملیاتی، بازده سهام و بازده شرکت) با ریسک تقلب است. فرج زاده دهکردی و آقایی(۱۳۹۴) رابطه بین تقلب در گزارشگری مالی و سیاست های تقسیم سود شرکت ها رابررسی کردند. ودر تحقیق خود بر شرایطی تمرکز کردند که بر اساس انگیزه های مدیریت در به کارگیری اقلام تعهدی اختیاری، می توان تجدید ارائه صورت های مالی را در دو گرو ه متقلبانه و غیرمتقلبانه، طبقه بندی نمود. نتایج به دست آمده نشان داد، شرکت هایی که سود تقسیم می کنند با احتمال کمتری مرتکب گزارشگری مالی متقلبانه می شوند.

ابراهیمی و همکاران(۱۳۹۵) در پژوهشی با عنوان «شناسایی پارامترهای تأثیرگذار بر تقلب در حوزه مالی» با استفاده از داده های جمع آوری شده از ۳۸۸ پرسشنامه و به کارگیری تاپسیس فازی به الویت بندی این پارامترها پرداختند. براساس نتایج حاصله، عدم استقلال واقعی بانک مرکزی، نبود سیستم نظارتی کارآمد و شفاف در دستگاه نظارتی و بازرسی مالی و چندین عامل دیگر به عنوان عوامل اصلی شناسایی شده است. سجادی وکاظمی(۱۳۹۵) در پژوهشی با عنوان«الگوی جامع گزارشگری مالی متقلبانه در ایران به روش نظریه‌پردازی زمینه بنیان» الگوی

جامع تقلب در صورت‌های مالی را در بستر فرهنگی، اقتصادی وحقوقی کشور را ارایه نمایند. جامعه آماری پژوهش خبرگان صاحب نظر در خصوص پدیده صورت‌های مالی متقلبانه هستند که با توجه به هدف پژوهش، از روش نمونه‌گیری گلوله‌برفی یا زنجیره‌ای برای مصاحبه انتخاب شده‌اند. و نتایج پژوهش آنان نشان می‌دهد که انگیزه پاداش مدیران، انگیزه سوءاستفاده از دارایی‌ها، هزینه‌های سیاسی، مقاصد مالیاتی و تحصیل شرکت توسط مدیران نیز برگزارشگری مالی متقلبانه موثراند. طرح‌های تقلب در گزارشگری مالی در بستر فرهنگ عمومی، نظام قانونی و استاندارد حسابداری کشور به عنوان شرایط زمینه‌ای و نظام راهبری شرکتی،کنترل داخلی و کیفیت حسابرسی به عنوان شرایط مداخله‌گر متولد می‌شوند. مهدوی و قهرمانی(۱۳۹۶) در پژوهشی با عنوان «ارائه الگویی برای کشف تقلب به وسیله حسابرسان با استفاده از شکبه عصبی مصنوعی» به این نتیجه رسیدند که الگوی شبکه عصبی طراحی شده با ۹ نورون در لایه پنهان دارای دقت ۸۶.۹ درصد توانایی شناسایی شرکت های متقلب و غیر متقلب را دارد. خواجوی و ابراهیمی(۱۳۹۶) در پژوهشی با عنوان «مدل سازی متغیرهای اثرگذار بر کشف تقلب در صورت های مالی با استفاده از الگوریتم های داده کاوی» به بررسی این موضوع پرداختند که آیا می توان از طریق شناسایی و انتخاب متغیرهای اثرگذار در کشف تقلب در صورت های مالی و با به کار گیری الگوریتم های داده کاوی مدلی برای کشف تقلب در صورت های مالی شرکت های پذیرفته شده در بورس اوراق بهادار تهران ارائه کرد؟ آنها برای این منظور از الگوریتم های طبقه بندی شبکه عصبی مصنوعی، شبکه بیزین و الگوریتم جنگل تصادفی و از متغیرهای پیش بینی، عوامل مالی و غیر مالی خطر تقلب مرتبط با گزارشگری مالی متقلبانه استفاده کردند. یافته های پژوهش بیانگر وجود شواهدی دال بر عملکرد مناسب مدل های پیشنهادی و برتری الگوی جنگل تصادفی و شبکه بیزین برای پیش بینی تقلب در صورت های مالی است. نتایج حاصل از انتخاب ویژگی به روش مبتنی بر همبستگی حاکی از سودمندی متغیرهای نسبت پوشش بهره، نسبت حسابهای دریافتنی به کل دارایی ها، نسبت موجودی کالا به فروش خالص، نسبت نقدی، لگاریتم طبیعی فروش، نسبت سود خالص به فروش و نسبت جمع دارایی های جاری به کل دارایی ها برای کشف تقلب بود.

اعتمادی و عبدلی(۱۳۹۶) در پژوهشی با عنوان «کیفیت حسابرسی و تقلب در صورت های مالی» به این نتیجه رسیدند که رابطه منفی و معناداری بین اندازه حســابرس، دوره تصدی حسابرس، تخصص حسابرس در صنعت، مؤسسه حسابرسی متخصص در صنعت با طول دوره تصدی طولانی و امتیاز وضعیت کنترل و کیفیت حسابرسی با تقلب در صورت های مالی وجود دارد. نتایج به طور کلی بیانگر این است که هر چه در شرکت ها، کیفیت حسابرسی بالاتر باشد ارتکاب شرکت‌ها به تقلب کمتر است. خواجوی و ابراهیمی(۱۳۹۶) در پژوهشی با عنوان «ارائۀ یک رویکرد محاسباتی نوین برای پیش بینی تقلب در صورت های مالی، با استفاده از شیوه های خوشه بندی و طبقه بندی (شواهدی از شرکت های پذیرفته شده در بورس اوراق بهادار تهران)» به بررسی این مسئله پرداختند که آیا می توان از طریق شناسایی عوامل مرتبط با تقلب در صورت های مالی و با به کارگیری شیوه های داده کاوی، مدلی برای کشف تقلب در صورت های مالی شرکت های پذیرفته شده در بورس اوراق بهادار تهران ارائه کرد؟ برای پاسخ گویی به این سؤال از ۱۹ علائم خطر اشاره شده در استاندارد حسابرسی ۲۴۰ به همراه شیوه های داده کاوی تحلیل مؤلفه های اساسی و خوشه بندی، برای تعیین شرکت های متقلب استفاده شد؛ سپس به

منظور ارائه مدلی برای پیش بینی صورت های مالی متقلبانه، از ۴۰ متغیر مالی و غیر مالی به همراه شیوه های درخت تصمیم، ماشین بردار پیشتیبان و روش بوستینگ استفاده شد. یافته های پژوهش بیان گر وجود شواهدی دال بر عملکرد مناسب مدل های پیشنهادی برای پیش بینی تقلب در صورت های مالی است.

میرفندرسکی و همکاران (۱۳۹۷) در پژوهشی با عنوان" رابطه ی بین دوره تصدی حسابرس و خطر تقلب در گزارشگری مالی با کنترل اثر شرایط نااطمینانی محیطی"به منظور قابل اتکاتر شدن نتایج پژوهش، اثر نااطمینانی محیطی بر رابطه ی بین متغیرهای مورد مطالعه کنترل گردیده است. نمونه ی پژوهش مشتمل بر ۸۶ شرکت از شـــرکتهای پذیرفته شده در بورس اوراق بهادار تهران در بازه ی زمانی ســـالهای ۱۳۹۰ الـــی ۱۳۹۴ جمع آوری گردید. برای آزمون فرضیه ی پژوهش از الگوی الجیت استفاده شد نهایتا به عدم وجود رابطه ی معنا دار بین دوره ی تصدی حسابرس و خطر تقلب در گزارشگری مالی دست یافتند. حاجی حیدری و رحیمیان (۱۳۹۸)در پژوهشی با عنوان" کشف تقلب با استفاده از مدل تعدیل شده بنیش و نسبت‌های مالی " پس از اجرای مدل رگرسیون در سه مرحله، نتایج نشان می‌دهند، نسبت فروش به مجموع دارایی‌ها و نسبت حقوق صاحبان سهام به مجموع دارایی‌ها دو نسبت مالی حساس به تقلب هستند. این مدل در طبقه بندی نمونه موردنظر در این تحقیق از نرخ دقت کلی ۶۹/۱ درصد برخورداراست. در نتیجه این مدل نقش اثربخشی درکشف تقلب صورت‌های مالی داشته است. ملانظری و شمس (۱۳۹۸) در پژوهشی با عنوان " مدل عوامل مؤثر بر قضاوت و تصمیم‌گیری در خصوص تقلب مرتبط با دارایی‌ها" بیان کردند سازگاری، وجدان، برون‌گرایی و گشودگی به‌تجربه رابطه معناداری با یادگیری هدف‌گرا دارند اما میان گشودگی به تجربه و عملکرد هدف‌گرا رابطه معناداری یافت نشد. گرایش هدف‌گرا اثر معناداری بر دانش داشته اما میان عملکرد هدف‌گرا و تجربه رابطه معناداری یافت نشد. همچنین دانش و تجربه اثر معناداری بر شناسایی عوامل ریسک دارند که منجر به تشخیص امکان رخداد تقلب می‌شود. نهایتاً نتایج حاصل از مدل آزمون‌شده نشان می‌دهد که تصمیم افراد برای گزارش موارد مستعد تقلب تحت تاثیر رابطه معنادار با تشخیص امکان رخداد تقلب، اخلاق و انگیزش می‌باشد.

فلاح حمیدی و همکاران (۱۳۹۹) در پژوهشی با عنوان «قدرت و بیش اطمینانی مدیران و گزارشگری متقلبانه» نشان دادند که بین قدرت مدیران و گزارشگری متقلبانه رابطه معناداری وجود ندارد. با این حال بین بیش اطمینانی مدیران و گزارشگری متقلبانه رابطه معناداری مشاهده شد. رضائی و همکاران(۱۳۹۹) در پژوهشی با عنوان «کشف تقلب صورت های مالی با توجه به گزارش حسابرسی صورت های مالی» از طریق بررسی اطلاعات ۱۶۴ شرکت پذیرفته شده در بورس اوراق بهادار تهران در طی دوره زمانی ۱۳۹۳ تا ۱۳۹۶ و با بکارگیری الگوریتم های یادگیری ماشین شامل؛ شبکه های بیزین، درخت تصمیم، شبکه های عصبی، ماشین بردار پشتیبان و روش ترکیبی، نتایج نشان دادند تمامی تکنیک ها قابلیت کشف تقلب صورت های مالی را در سطح نسبتا بالایی دارند و تکنیک پیشنهادی ترکیبی با میزان نرخ پیش بینی ٪۹۶.۲ دارای دقت و توان ارزیابی بالاتری نسبت به سایر تکنیک ها است. احمدی و همکاران (۱۳۹۹) در پژوهشی با عنوان «ارائه مدلی برای پیش بینی صورت‌های مالی متقلبانه و مقایسه صورت ها و نسبت های مالی با قانون بنفورد» براساس نمونه آماری شامل ۴۱۰ سال-شرکت متقلب و ۴۱۰ سال-شرکت غیر متقلب و با استفاده از روش رگرسیون لجستیک نشان دادند که با توجه به نرخ دقت ۶۴.۶ درصدی، این مدل نقش

اثر بخشی در کشف تقلب صورت های مالی دارد. همچنین نتایج آزمونT و آزمون لون در ۳۵ متغیر مستقل بررسی شده نشان داد که در ۲۰ متغیر، تفاوت معناداری در دو گروه متقلب و غیر متقلب وجود دارد. به علاوه تطابق پذیری و انحراف از قانون بنفورد در چهار حالت مختلف بررسی شد و نتایجی بدین شرح حاصل گردید که؛ توجه به ارقام صورت سود و زیان و ترازنامه در شرکت های غیر متقلب نشان داد توزیع بنفورد شرکت های غیر متقلب را به درستی تشخیص داده ولی شرکت های متقلب را به صورت نادرست غیر متقلب تشخیص داده است. توجه به نسبت مالی کل دارایی ها به فروش و دوره پرداخت حسابهای پرداختنی در شرکت های متقلب نشان داد که توزیع بنفورد این شرکتها را متقلب ارزیابی و به درستی دسته بندی کرده است ولی شرکت های غیرمتقلب را به صورت نادرست متقلب ارزیابی کرده است.

پسندیده فرد و همکاران (۱۳۹۹) در پژوهشی با عنوان «شناسایی عوامل موثر بر گزارشگری مالی متقلبانه و نادرست با استفاده از روش فراترکیب» با رویکرد پژوهش کیفی و ابزار فراترکیب (متاسنتز) که شامل گام‌های هفت‌گانه ای است، به ارزیابی و تحلیل نظام مند ۳۱۱ مورد از یافته های پژوهش‌های پیشین پرداخته شده است. در انتها نظر ۱۸ نفر از خبرگان و اساتید، به وسیله پرسشنامه در سال ۱۳۹۸ جمع‌آوری شده و با استفاده از روش کمی آنتروپی شانون، براساس رویکرد تحلیل محتوا به تعیین ضریب اثر عوامل شناسایی شده، پرداخته شد. در نهایت عواملی که بیش‌ترین تاثیر را بر "گزارشگری مالی متقلبانه و نادرست" دارند، تعیین گردید. یوخنه القیانی و همکاران (۱۳۹۹) در پژوهشی با عنوان «تبیین الگوی تقلب مالیاتی در ایران مبتنی بر رویکرد آمیخته (واکاوی کنش مالیاتی متقلبانه اشخاص حقوقی)» به این نتیجه رسیدند که در بخش کیفی، رویکرد مالیاتی متقلبانه به عنوان مقوله هسته ای تعیین و ابعاد مختلف مدل، حاصل گشت: کنش جبرمحور و کنش گزینشی به عنوان علل شکل گیری رویکرد مالیاتی متقلبانه تبیین شد. شرایط بستر شامل ۷ محتوا، راهبردهای متقلبانه در ۸ گروه راهبردهای متقلبانه خاص و ۹ گروه راهبرد متقلبانه عمومی، شرایط مداخله گر تشدیدگر و تحدیدگر در سه حوزه سازمان مالیاتی، خارج از دستگاه مالیاتی و حوزه مشترک، همراه با ۵ سطح پیامدهای اقتصادی و تبعات اجتماعی-سیاسی تبیین و تحلیل گشت. نتایج تحلیل کمی نشان داد که کنش جبرمحور و کنش گزینشی تاثیر مستقیم و معنادار، بایستگی نظام مالیاتی، بایستگی نظامات اجتماعی و بایستگی عملکرد مشترک، اثر معکوس و معنا دار و عوامل زمینه ای اثر مستقیم و غیرمستقیم معنادار بر رویکرد مالیاتی متقلبانه شرکت ها دارند. ملکی کاکلر و همکاران (۱۴۰۰) در پژوهشی با عنوان «کارایی مدل های آماری والگوهای یادگیری ماشین در پیش بینی گزارشگری مالی متقلبانه» با استفاده از ۲۰ متغیر در قالب الگوی پنج ضلعی تقلب با تاکید بر ساختار کنترل های داخلی در ۱۶۶ شرکت های فعال در بورس اوراق بهادار تهران طی سال های ۱۳۸۸ الی ۱۳۹۷ و مقایسه بین مدل های مورد بررسی، باکمک آزمون مقایسه نسبت ها، نتایج نشان می دهد که به لحاظ آماری مدل های یادگیری ماشین در پیش بینی گزارشگری مالی متقلبانه نسبت به مدل های آماری، کارایی و دقتب یشتری دارند. ترکیب الگوریتم درخت تصمیم گیری CHAID، C5 و C&R بالاترین دقت در پیش بینی گزارشگری مالی متقلبانه را با دقت بالای ۹۲/۶۱ درصد در پیش بینی تقلب نشان می دهد.

سیاهکارزاده (۱۴۰۰) در پژوهشی با عنوان «پیش‌بینی بازارهای مالی با استفاده از یادگیری عمیق» یک مدل از شبکه‌های عصبی با استفاده از معماری‌های CNN و LSTM و با الهام از شبکه عصبی چند فیلتری (MFNN) ارائه داد که تفاوت عمده آن با سایر مدل‌های ارائه شده توانایی پیش‌بینی در قالب‌های زمانی کوتاه مدت می‌باشد و همچنین با بررسی روش‌های مختلف برای استخراج و تولید ویژگی، یک روش جدید، به‌طور خاص برای استخراج ویژگی در نمونه‌های سری‌زمانی مالی و پیش‌بینی روند قیمت پیشنهاد کرد. در این روش از ترکیب انواع شاخص‌ها، استراتژی‌های جدیدی تولید شد که از آن در تولید ویژگی استفاده کرده، با مقایسه دادگان پیشنهادی با سایر دادگان در شرایط یکسان شاهد افزایش دقت پیش‌بینی حداقل ۱.۸٪ و حداکثر ۶٪ بودیم. در ادامه با معرفی روشی جدید در هدف گذاری داده‌های مالی و مقایسه با سایر روش‌ها، مراحل پیش‌پردازش بر روی دادگان سری زمانی مالی را به اتمام رساند. بعد از مقایسه روش‌های برچسب‌گذاری شاهد افزایش دقت حداقل ۱.۲۵ و حداکثر ۵ درصدی هستیم. همچنین در ادامه در مدل ارائه شده از ترکیب هر دو معماری شبکه عصبی CNN و LSTM استفاده شده است. نویسنده روش خود را برای پیش‌بینی مهمترین و پر حجم‌ترین بازارمالی دنیا یعنی بازار تبادل ارزهای خارجی اعمال کرد. مدل پیشنهادی دقت ۷۸.۴۲٪ را ارائه می‌دهد که نسبت به بهترین عملکرد در مدل های معرفی شده رشد ۳ درصدی دارد. یوخنه القیانی و همکاران (۱۴۰۰) در پژوهشی با عنوان «تبیین گزارشگری مالی- مالیاتی متقلبانه شرکت ها: رویکرد ترکیبی داده کاوی کلاسیک، ANFIS و الگوریتم های فراابتکاری» ابتدا اثرگذاری نشانگرهای کیفی و کمی برگرفته از گزارش های مالی به روش کلاسیک بررسی، پالایش و سپس جهت تبیین مدل در سیستم استنتاج فازی- عصبی تطبیقی (ANFIS) به کارگرفته شد. مدل مذکور در مرحله بهینه سازی و با هدف بیشینه سازی قدرت تشخیص، با الگوریتم های بهینه سازی فراابتکاری شامل الگوریتم ژنتیک، الگوریتم بهینه سازی ازدحام ذرات و الگوریتم تکامل تفاضلی ترکیب و توسعه یافت. یافته های پژوهش نشان می دهد که از حیث قدرت پیش بینی در میان الگوریتم های بهینه سازی به کارگرفته شده در پژوهش، الگوریتم ازدحام ذرات با بیشترین درصد پیش بینی صحیح، بهینه ترین مدل را حاصل نموده و در بررسی توسط داده های آزمایشی و آموزشی کاراترین الگوریتم است. نتایج حاصل از پژوهش انجام شده حاکی از این است که به کارگیری الگوریتم های بهینه سازی مختلف در رویکرد داده کاوی، سبب افزایش قدرت پیش بینی صحیح مدل شناسایی گزارشگری مالی- مالیاتی متقلبانه می گردد. ملکی کاکلر و همکاران (۱۴۰۰) در پژوهشی با عنوان «ارائه مدل توسعه یافته ی پنج ضلعی تقلب در گزارشگری مالی با تاکید بر ساختار کنترلهای داخلی» به این نتیجه رسیدند که متغیرهای فشار، فرصت، توجیه، قابلیت و ساختار کنترل های داخلی دارای تاثیر معناداری بر تقلب گزارشگری مالی متقلبانه هستند. از سوی دیگر متغیر تکبر تاثیر معنادار در گزارشگری مالی متقلبانه ندارد.

خلیلی ثمرین و همکاران (۱۴۰۰) در پژوهشی با عنوان «رتبه‌بندی عوامل درون‌سازمانی و برون سازمانی موثر بر گزارشگری مالی متقلبانه با استفاده از فرآیند تحلیل سلسله مراتبی» نشان دادند که عوامل برون سازمانی در ایجاد گزارشگری متقلبانه با امتیاز ۰.۲۴۳ در اولویت اول بوده و عوامل درون سازمانی با امتیاز ۰.۲۱۴ در اولویت بعد قرار گرفت. همچنین با توجه به نتایج پژوهش، از عوامل برون سازمانی به ترتیب عوامل مرتبط با ویژگی های حسابرسی مستقل، عوامل فرهنگی، عوامل قانونی و نظارتی خارج از سازمان و از عوامل درون سازمانی به ترتیب

ویژگی های حرفه ای و ساختار مدیریت، ویژگی های رفتاری مدیریت، ویژگی های رفتاری و اخلاقی، ویژگی های مرتبط با خطاهای سیستمی و انسانی دارای بیشترین اهمیت بوده است. رضائی و همکاران (۱۴۰۰) در پژوهشی با عنوان «پیش بینی تقلب صورت های مالی با استفاده از رویکرد کریسپ (CRISP)» متغیرهای مستقل تاثیر گذار بر تقلب در این پژوهش در برگیرنده ۴۰ متغیر مالی و غیر مالی می باشد که براساس مبانی نظری و پیشینه پژوهش انتخاب شده اند. در نهایت داده های مربوط به متغیرها که به روش کتابخانه ای جمعآوری شد هاند، براساس رویکرد کریسپ، جهت تعیین وزن و ویژگی متغیرهای بااهمیت به مدل آنتروپی شانون و به منظور پیشبینی تقلب به چهار تکنیک برتر از بین تکنیکهای هوش مصنوعی داده شده است، که این تکنیک ها شامل؛ درخت تصمیم، شبکه های عصبی، ماشین بردار پشتیبان و روش ترکیبی آدابوست ماشین بردار پشتیبان می باشد. با استفاده از آنتروپی شانون از بین ۴۰ متغیر پژوهش، ۲۷ متغیر برتر براساس ویژگی سود اطلاعاتی، مشخص گردید، که متغیر نسبت سود انباشته به فروش به عنوان با اهمیت ترین متغیر در زمینه پیش بینی تقلب صورت های مالی شناسایی شده است .پس از بکارگیری رویکرد کریسپ، نتایج نشان داد تمامی تکنیک ها قابلیت کشف تقلب صورت های مالی را در سطح نسبتا بالایی دارند و تکنیک پیشنهادی آدابوست ماشین بردار پشتیبان در مرحله آموزش با نرخ دقت ۸۱.۶۹٪ دارای دقت و توان ارزیابی باالتری نسبت به سایر تکنیک ها بوده و این تکنیک در مرحله آزمایش ۸۲٪ صورت های مالی متقلبانه و غیرمتقلبانه سال ۱۳۹۶ را بدرستی تشخیص داد. بهشتی مسئله گو و همکاران (۱۴۰۱) در پژوهشی با عنوان «یادگیری عمیق برای پیش‌بینی بازار سهام با استفاده از اطلاعات عددی و متنی (رویکرد الگوریتم حافظه کوتاه مدت ماندگار LSTM) » از داده‌های سهام شاخص داوجونز و داده‌های خبری کانال ردیت استفاده شده است. از داده‌های سهام ویژگی‌های مبتنی بر شاخص فنی استخراج می‌شوند و داده‌های خبری توسط روش کوله‌کلمات به بردار ویژگی تبدیل می‌شوند و به شبکه حافظه کوتاه‌مدت ماندگار (LSTM) برای پیش‌بینی داده می‌شوند.از دقت به‌عنوان معیار ارزیابی عملکرد استفاده شده و آزمایش‌هایی بر روی دو مجموعه داده فقط عددی و فقط متنی برای ارزیابی استفاده همزمان دو منبع اطلاعاتی انجام پذیرفته است. همچنین ازسه شبکه،SVM، MLP و RNNبرای ارزیابی مدل استفاده شده است. نتایج نشان می‌دهد که مدل LSTM بالاترین دقت پیش‌بینی ۶۹.۱۹٪ را با استفاده از اخبار و داده‌های مالی به دست آورده است. داده‌های خبری با دقت۶۵.۶۲٪ و داده‌های عددی با دقت۵۱.۸۹٪ می‌باشند.همچنین مدل LSTM در مقایسه با شبکه‌های عصبی SVM و MLP و RNN از عملکرد بهتری برخوردار می‌باشد.

شکوری و همکاران (۱۴۰۱) در پژوهشی با عنوان «تبیین مدل پنتاگون تقلب و ارائه مدل جامع گزارشگری مالی متقلبانه» در بررسی جداگانه شاخص های مدل پنتاگون، نتایج پژوهش بیانگر آن است که از بین شاخص های فشار، فشارهای بیرونی، که موسسات مالی و اعتباری، بر شرکت ها اعمال می کنند، تاثیر بالایی بر گزارشگری مالی متقلبانه داشته است. در بررسی شاخص فرصت، با استفاده از ماهیت صنعت، شواهد بیانگر آن است که این شاخص دارای ارتباط معکوس با گزارشگری مالی متقلبانه است و هر چه این شاخص کاهش یابد، احتمال وقوع تقلب در صورتهای مالی، بیشتر می شود. در بررسی شاخص های توجیه، تغییر حسابرسان مستقل، از اهمیت بالاتری برخوردار بوده و در حقیقت تغییر زیاد حسابرسان مستقل، امکان بیشتری برای بروز تقلب را فراهم می کند. در

بررسی شاخص های استعداد، افزایش مکرر در حق الزحمه های غیرعادی حسابرسی، ممکن است حسابرسان مستقل را، از افشای موارد تقلب کشف شده، در گزارشگری مالی، باز دارد و صورتهای مالی را مستعد بروز رفتارهای متقلبانه، نماید و در نهایت از بین شاخص های خودشیفتگی، تغییرات شدید در قیمت سهام، از اهمیت بالاتری در زمینه ارتکاب تقلب در گزارشگری مالی برخوردار بوده و هر چه میزان تغییرات شدید در قیمت سهام، بیشتر باشد، احتمال بروز تقلب در گزارشگری مالی، بیشتر می شود. در بررسی کلی شاخص ها نیز، شاخص های فشار، بیشترین اثر را، بر گزارشگری مالی متقلبانه داشته است. کاظمی و پیری(۱۴۰۱) در پژوهشی با عنوان «پیش بینی طرح تقلب در گزارشگری مالی با استفاده از رویکرد یادگیری ماشین در فضای چند کلاسه» به این نتیجه رسیدند که تفاوت معنادار در عملکرد الگوهای یادگیری ماشین در فضای چند کلاسه وجود دارد و روش ماشین بردار پشتیبان نسبت به سایر روش ها عملکرد بهتری دارد. با تقلیل فضای مسیله به دسته بندی دو کلاسه تفاوت معنادار در عملکرد الگوهای یادگیری ماشین در تشخیص گزارش های مالی مشکوک به "بیش نمایی دارایی، کم نمایی بدهی و هزینه"، "بیش نمایی دارایی و کم نمایی هزینه" و "کم نمایی هزینه و بدهی" تایید نشد. با این حال، عملکرد ماشین بردار پشتیبان بر عملکرد روش رگرسیون لجستیک و درخت تصمیم در پیش بینی گزارش های مالی مشکوک به "بیش نمایی دارایی و درآمد" ارجح است. صفی خانی و همکاران (۱۴۰۲) در پژوهشی با عنوان «پیشایندهای گزارشگری مالی متقلبانه در شرکت های دارای بحران مالی» به منظور تفکیک آن به دو گروه شرکتهای دارای گزارشگری مالی متقلبانه و شرکتهای فاقد گزارشگری مالی متقلبانه، از بندهای گزارش حسابرسی استفاده کردند. نتایج حاصل از آزمون فرضیه های پژوهش نشان داد: ۱) فراوانی شرکت های دارای بحران مالی که به گزارشگری مالی متقلبانه روی می آورند، به طور معناداری بیشتر از شرکت های دارای بحران مالی که به گزارشگری مالی متقلبانه روی نمی آورند، است. ۲) در شرکت های دارای بحران مالی و دارای گزارشگری مالی متقلبانه در دوره مالی قبل و دوره مالی جاری، میزان عدم تقارن اطلاعاتی میان مدیران و ذینفعان، به طور معناداری بیشتر از شرکت های دارای بحران مالی و فاقد گزارشگری مالی متقلبانه است. ۳) در شرکت های دارای بحران مالی و دارای گزارشگری مالی متقلبانه در دوره مالی قبل و دوره مالی جاری، فراوانی شرکت های دارای بیش سرمایه گذاری، به طور معناداری بیشتر از شرکت های دارای بحران مالی و فاقد گزارشگری مالی متقلبانه است. ۴) بین شرکت های دارای بحران مالی و دارای گزارشگری مالی متقلبانه و شرکت های دارای بحران مالی و فاقد گزارشگری مالی متقلبانه در دوره مالی بعد، در زمینه آسیب به حسن شهرت شرکت، تفاوت معناداری وجود ندارد. بابانژاد باقری و همکاران (۱۴۰۲) در پژوهشی با عنوان «پیش بینی ارزش شرکت مبتنی بر روش‌های یادگیری عمیق» نتایج حاصل از بررسی داده های گردآوری شده با استفاده از تکنیک های یادگیری عمیق، بیانگر آن بود که مدل ترکیبی با مقدار خطای RMSE کمتری نسبت به مدل GRU ارزش شرکت را پیش بینی کرده است.

۳. روش پژوهش

۳-۱. نحوه جمع آوری داده ها

این پژوهش از لحاظ هدف کاربردی است. طرح پژوهش آن از نوع پژوهش های شبه تجربی و پس رویدادی است و با استفاده از اطلاعات تاریخی انجام می شود. برای گردآوری اطلاعات مبانی نظری پژوهش از نشریات، کتب و همچنین پایگاه های اطلاعاتی در دسترس استفاده شده است. همچنین، داده های مورد نیاز برای تجزیه و تحلیل، از نرم افزار "ره آورد نوین" و سایت های اینترنتی "مدیریت پژوهش، توسعه و مطالعات اسلامی سازمان بورس اوراق بهادار"، کدال، بانک مرکزی و مرکز آمار ایران گردآوری شده است.

۳-۲. جامعه آماری، نمونه آماری و بازه ی زمانی پژوهش

جامعه آماری پژوهش حاضر، کلیه شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران است. بازه ی زمانی نیز یک دوره زمانی ۱۲ ساله براساس صورتهای مالی سال های ۱۳۹۱ تا ۱۴۰۲ شرکت های نمونه می باشد.

در پژوهش حاضر برای تعیین نمونه آماری، از روش غربالگری استفاده شده است. بدین منظور آن دسته از شرکت های جامعه آماری که شرایط زیر را دارا باشند به عنوان نمونه آماری انتخاب و مابقی حذف می شوند.

۱) *سال مالی شرکت منتهی به تاریخ پایان اسفند ماه هر سال باشد.*
۲) *شرکت طی دوره مورد بررسی تغییر سال مالی نداده باشد.*
۳) *شرکت های تحت بررسی جزء شرکتهای سرمایه گذاری، هلدینگ ، واسطه گری مالی و بیمه نباشند.*
۴) *اطلاعات و داده های آنها در دسترس باشد.*

با توجه به شرایط و محدودیت‌های فوق، از بین شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران در مجموع ۱۵۰ شرکت به عنوان نمونه آماری پژوهش انتخاب شده است.

جدول ۱: نحوه نمونه گیری

تعداد کل شرکت های پذیرفته شده در بورس اوراق بهادار تهران	577
کسر می شود:	
تعداد شرکت هایی که سال مالی آنها منتهی به پایان اسفند نمی باشد و یا طی دوره تحقیق تغییر سال مالی داده باشند.	(93)
تعداد شرکت هایی که در گروه شرکت های هلدینگ، سرمایه گذاری و یا واسطه گری های مالی بوده اند.	(124)
تعداد شرکت هایی که در قلمرو زمانی ۱۴۰۲-۱۳۹۱ اطلاعات کامل آن ها در دسترس نمی باشد.	(99)
تعداد شرکت های که در قلمرو زمانی ۱۴۰۲-۱۳۹۱ بیش از سه ماه توقف معاملاتی داشته اند.	(111)
تعداد شرکت های نمونه	150

منبع: یافته های پژوهشگر

۳-۳. روش اندازه گیری متغیرهای پژوهش

متغیر وابسته (پاسخ):

متغیر وابسته در پژوهش حاضر صورت های مالی متقلبانه می باشد که با استفاده از شاخص F_SCORE تدوین شده توسط دیچو و همکاران[1] (۲۰۱۱) اندازه گیری می شود که به صورت زیر می باشد:

رابطه (۱):

$$Predicted\ Value = -7.893 + 0.790\times(rsst\_acc) + 2.518\times(ch\_rec) + 1.191\times(ch\_inv) + 1.979\times(soft\_assets) + 0.171\times(ch\_cs) - 0.932\times(ch\_roa) + 1.029\times(issue)$$

که در آن:

rsst_acc: برابر است با (Δfin + Δwc + Δnco) تقسیم بر میانگین کل دارایی ها. wc برابر دارایی های جاری منهای وجه نقد و سرمایه گذاری های کوتاه مدت منهای بدهی های جاری است. nco برابر مجموع دارایی ها منهای دارایی جاری منهای سرمایه گذاری ها منهای مجموع بدهی ها به کسر بدهی های جاری و بلندمدت است. fin برابر سرمایه گذاری کوتاه مدت به علاوه ی سرمایه گذاری بلندمدت منهای مجموع بدهی های بلندمدت، بدهی موجود در بدهی های جاری و سهام ممتاز است.

ch_rec: برابر با تغییر در حساب های دریافتنی در طول دوره ی جاری تقسیم بر میانگین دارایی ها است.

ch_inv: برابر با تغییر در موجودیها در طول دوره ی جاری تقسیم بر میانگین دارایی ها است.

soft_assets: برابر با مجموع دارایی ها به کسر اموال، ماشین آلات و تجهیزات و وجه نقد و معادل آن تقسیم بر مجموع دارایی ها است.

ch_cs: برابر با درصد تغییر در فروش نقدی است، که از طریق فروش دوره ی جاری منهای تغییر در حساب های دریافتنی محاسبه می شود.

ch_roa: برابر با تغییر در بازده داراییها است، که از طریق تقسیم سود خالص بر میانگین دارایی ها بدست می آید.

Issue: متغیری ساختگی است، به این ترتیب که برای شرکتی که اوراق بدهی یا اوراق مالکیت منتشر کرده باشد عدد یک و در غیر این صورت صفر می باشد.

برای محاسبه F_SCORE، احتمال پیش بینی از طریق تقسیم $e^{PV}/(1+e^{PV})$ بر احتمال غیر شرطی ارائه نادرست صورت های مالی (۰.۰۰۳۷) بدست می آید. که در آن PV ارزش پیش بینی بدست آمده از رابطه (۱) می باشد. براساس پژوهش دیچو و همکاران (۲۰۱۱)، در این پژوهش مشاهدات دارای F_SCORE بالاتر از ۱.۸۵ به عنوان شرکت های با ریسک بالای صورت های مالی متقلبانه شناسایی می شوند.

[1] Dechow et al.

متغیرهای مستقل (پیش بین):

به پیروی از پژوهش های پیشین (علی و همکاران، ۲۰۲۳؛ هیوگو و شنگ لینگ[1]، ۲۰۲۲؛ لونگ جان[2]، ۲۰۲۱؛ کات سیس و همکاران[3]، ۲۰۱۲؛ پرلوس[4]، ۲۰۱۱؛ کی روکس و همکاران[5]، ۲۰۱۱؛ رضائی و همکاران، ۱۴۰۰؛ خواجوی و ابراهیمی، ۱۳۹۶؛ اعتمادی و زلقی، ۱۳۹۲؛ صفرزاده، ۱۳۸۹) متغیرهای مستقل پژوهش حاضر به شرح جدول زیر ارائه می گردد:

جدول ۲: نحوه اندازه متغیرهای مستقل پژوهش

ردیف	متغیر مستقل	نماد	نحوه اندازه گیری
۱	حاشیه سود ناخالص	GPM	سود ناخالص تقسیم بر فروش
۲	حاشیه سود خالص	NPM	سود خالص تقسیم بر فروش
۳	حاشیه سود عملیاتی	OPM	سود عملیاتی تقسیم بر فروش
۴	بازده دارایی ها	ROA	سود خالص تقسیم بر کل دارایی
۵	بازده حقوق صاحبان سهام	ROE	سود خالص تقسیم بر ارزش دفتری حقوق صاحبان سهام
۶	نسبت سود قبل از بهره و مالیات	EBIT/A	سود قبل از بهره و مالیاتی تقسیم بر کل دارایی
۷	گردش کل دارایی ها	ASTRN	فروش خالص تقسیم بر کل دارایی
۸	گردش دارایی های ثابت	FXASTRN	فروش خالص تقسیم بر دارایی های ثابت
۹	گردش موجودی کالا	INVTRN	بهای تمام شده کالای فروش رفته تقسیم بر متوسط موجودی کالا
۱۰	گردش حساب های دریافتنی	ACRECTRN	فروش خالص تقسیم بر متوسط حسابهای دریافتنی
۱۱	نسبت کل بدهی ها به کل دارایی ها	LIB/ASST	کل بدهی ها تقسیم بر کل دارایی ها
۱۲	نسبت کل بدهی ها به حقوق صاحبان سهام	LIB/EQT	کل بدهی ها تقسیم بر ارزش دفتری حقوق صاحبان سهام
۱۳	نسبت سود انباشته به کل دارایی ها	ACUM/ASST	سود(زیان) انباشته تقسیم بر کل دارایی
۱۴	نسبت سود انباشته به سرمایه	ACUM/EQT	سود(زیان) انباشته تقسیم بر سرمایه سهام عادی
۱۵	نسبت پوشش هزینه بهره	INTCOV	هزینه بهره تقسیم بر سود قبل از بهره و مالیات
۱۶	نسبت جاری	CURRAT	دارایی های جاری تقسیم بر بدهی های جاری
۱۷	نسبت آنی	QUICRAT	(وجه نقد + حسابهای دریافتنی + سرمایه گذاری کوتاه مدت) تقسیم بر بدهی های جاری

[1] Hugo and Sheng Ling
[2] Long John
[3] Cutsis et al
[4] Perlos
[5] Kay Roux et al

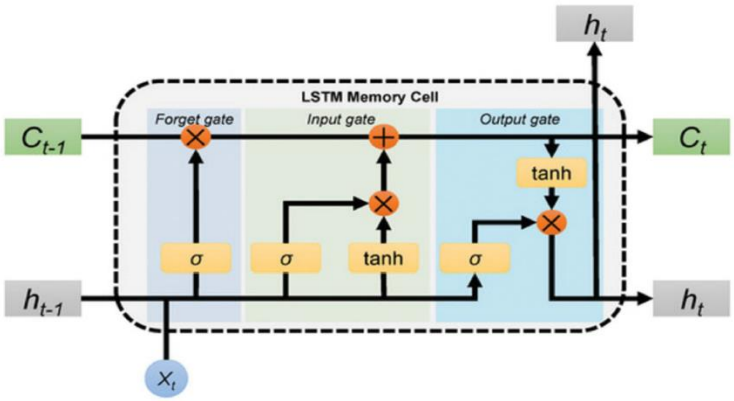
ردیف	متغیر مستقل	نماد	نحوه اندازه گیری
۱۸	نسبت وجه نقد آزاد به درآمد	FCF/REV	{سود خالص + هزینه استهلاک + هزینه مالی*(۱- مالیات) – سرمایه گذاری در دارایی ثابت – سرمایه گذاری در سرمایه در گردش} تقسیم بر فروش
۱۹	نسبت مخارج سرمایه ای به درآمد	CE/REV	مخارج سرمایه ای (تفاوت خالص ارزش دفتری دارایی های ثابت در ابتدا و پایان دوره به علاوه هزینه استهلاک) تقسیم بر فروش
۲۰	نسبت مخارج سرمایه ای به سود خالص	CE/NP	مخارج سرمایه ای تقسیم بر سود خالص
۲۱	نسبت وجه نقد عملیاتی به درآمد	OCF/REV	خالص جریان وجوه نقد عملیاتی حاصل از فعالیت های عملیاتی تقسیم بر فروش خالص
۲۲	نسبت وجه نقد عملیاتی به بدهی	OCF/LIB	خالص جریان وجوه نقد عملیاتی حاصل از فعالیت های عملیاتی تقسیم بر کل بدهی ها
۲۳	ارزش شرکت	EV	لگاریتم طبیعی (ارزش بازار سهام + ارزش دفتری بدهی ها – نقد و معادل نقد)
۲۴	ارزش دفتری شرکت	BV	لگاریتم طبیعی ارزش دفتری حقوق صاحبان سهام

منبع: یافته های پژوهشگر

۴-۳. الگوریتم های یادگیری عمیق مورد استفاده در پژوهش

الگوریتم حافظه‌ی طولانی کوتاه مدت(LSTM[1]): حافظه طولانی کوتاه مدت (LSTM) یک معماری شبکه عصبی بازگشتی است که در سال ۱۹۹۷ میلادی توسط سپ هوخرایتر و یورگن اشمیدهوبر ارائه شد، و بعداً در سال ۲۰۰۰ میلادی توسط فیلیکس ژرس و دیگران بهبود داده شد. در شبکه عصبی بازگشتی برخلاف شبکه عصبی پیشخور، ورودی می‌تواند به صورت دنباله (مثل صوت یا ویدیو) باشد. این ویژگی باعث شده شبکه‌های عصبی بازگشتی برای پردازش داده‌های زمانی بسیار مناسب باشند زیرا الگوها در داده‌های سری زمانی می‌توانند در فواصل مختلف واقع شوند. برای مثال، شبکه LSTM در حوزه‌هایی همچون تشخیص دست خط، ترجمه ماشینی، تشخیص گفتار، کنترل ربات‌ها و بازی‌های ویدیویی قابل استفاده است.

[1] Long short-term memory



شکل ۱: یک بلوک حافظه طولانی کوتاه مدت (فن و همکاران، ۲۰۲۰)

منبع: یافته های پژوهشگر

یک بلوک LSTM به طور کلی از وزن‌های محتوای سلول (c_t)، دروازه ورودی(i_t)، دروازه خروجی(o_t) و دروازه فراموشی(f_t) تشکیل شده است. محتوای سلول می‌تواند در بازه زمانی طولانی محتوای خود راه حفظ کند و در واقع اطلاعات پیشین را به خاطر بسپارد. دروازه‌ها نیز بردارهایی با مقادیر بین ۰ و ۱ هستند که چگونگی پیشروی اطلاعات قدیمی و اضافه شدن اطلاعات جدید را مشخص می‌کنند. به طور کلی ۱ به معنای عبور و ۰ به معنای حذف اطلاعات است. دروازه ورودی مشخص می‌کند چه بخش‌هایی از داده ورودی و به چه مقدار به محتوای سلول اضافه شوند. دروازه فراموشی مشخص می‌کند که چه بخش‌هایی از محتوای سلول از آن حذف شوند. دروازه خروجی نیز مشخص می‌کند محتوای وضعیت مخفی (h_t) حاوی چه بخشی از محتوای سلول باشد.

یکی از مشکلات اصلی شبکه‌های عصبی بازگشتی، مشتق ناپدیدشونده[1] و مشتق انفجاری[2] است. در بلوک LSTM به منظور حل این دو مشکل، راه‌هایی برای انتقال اطلاعات مهم پیشین در هنگام پیشروی ایجاد شده‌است. این مهم به طور کلی از طریق ضرب بردارهایی تحت عنوان دروازه، در ورودی‌های بلوک صورت می‌پذیرد.

ماشین بردار پشتیبان(SVM[3]): ماشین بردار پشتیبان (SVM) یکی از روش‌های یادگیری بانظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. این روش از روش‌های نسبتا جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی نشان داده است. مبنای کاری دسته بندی‌کننده SVM دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کنیم خطی را انتخاب کنیم که حاشیه اطمینان بیشتری داشته باشد. حل معادله پیدا کردن خط بهینه برای داده‌ها به وسیله روش‌های QP که روش‌های شناخته شده‌ای در حل مسائل محدودیت دار هستند صورت می‌گیرد.

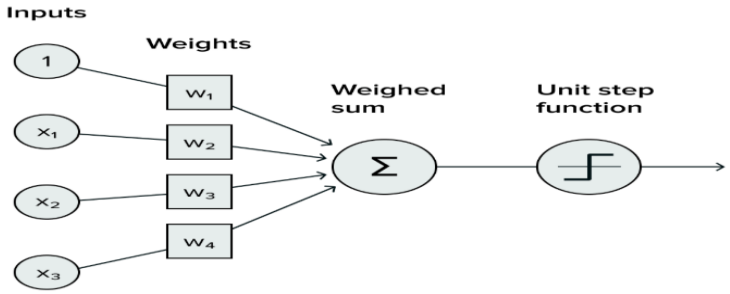
[1] vanishing gradient problem
[2] exploding gradient
[3] Support vector machines

قبل از تقسیم خطی برای اینکه ماشین بتواند داده‌های با پیچیدگی بالا را دسته‌بندی کند داده‌ها را با استفاده از تابعِ phi به فضای با ابعاد خیلی بالاتر می‌بریم. برای اینکه بتوانیم مسئله ابعاد خیلی بالا را با استفاده از این روش‌ها حل کنیم از قضیه دوگانی لاگرانژ برای تبدیل مسئله مینیمم‌سازی مورد نظر به فرم دوگانی آن که در آن به جای تابع پیچیده phi که ما را به فضایی با ابعاد بالا می‌برد، تابع ساده‌تری به نام تابع هسته که ضرب برداری تابع phi است ظاهر می‌شود، استفاده می‌کنیم. از توابع هسته مختلفی از جمله هسته‌های نمایی، چندجمله‌ای و سیگموید می‌توان استفاده نمود.

شبکه عصبی پیچشی(CNN[1]): شبکه‌ی عصبی پیچشی یا به اختصار CNN که به آن شبکه‌ی عصبی کانولوشنی نیز گفته می‌شود، نوعی از شبکه‌های عصبی است که عموماً برای یادگیری بر روی مجموعه داده‌های بصری (مانند تصاویر و عکس‌ها) استفاده می‌شود. از لحاظ مفهوم این شبکه‌ها مانند شبکه‌های عصبی ساده هستند یعنی از فازهای پیش‌خور[2] و پس‌انتشار خطا[3] استفاده می‌کنند ولی از لحاظ معماری تفاوت‌هایی با شبکه‌های عصبی ساده دارند. این شبکه‌ها در دسته‌ی یادگیری عمیق قرار می‌گیرند زیرا لایه‌های موجود در این شبکه‌ها، زیاد است.

شبکه عصبی پیچشی همانند سایر شبکه های عصبی از لایه‌های نورونی با وزن و بایاس با قابلیت یادگیری تشکیل شده است. لازم به ذکر است که در هر نورون اتفاقات زیر رخ می‌دهد:

۱) نورون مجموعه‌ای ورودی دریافت می‌کند.
۲) ضرب داخلی بین وزن‌های نورون و ورودی‌ها انجام می‌شود.
۳) حاصل با بایاس جمع می‌شود.
۴) درنهایت، از یک تابع غیرخطی[4] عبور داده می‌شود.



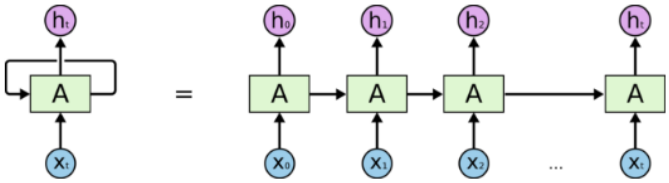
شکل ۲: یک بلوک شبکه عصبی پیچشی

منبع: یافته های پژوهشگر

[1] Convolutional Neural Networks
[2] feed forward
[3] back propagation of error
[4] activation function

فرآیند بالا لایه به لایه انجام می‌شود و درنهایت به لایه خروجی می‌رسیم. لایه خروجی، پیش‌بینی شبکه را تولید می‌کند.

شبکه عصبی بازگشتی (RNN[1]): شبکه عصبی بازگشتی (RNN) که به آن شبکه عصبی مکرر نیز گفته می شود، نوعی از شبکه عصبی مصنوعی است که در تشخیص گفتار، پردازش زبان طبیعی (NLP) و همچنین در پردازش داده های ترتیبی[2] استفاده می شود. بسیاری از شبکه های عمیق مانند CNN شبکه های پیش خور[3]هستند یعنی سیگنال در این شبکه ها فقط در یک جهت از لایه ورودی، به لایه های مخفی و سپس به لایه خروجی حرکت می کند و داده های قبلی به حافظه سپرده نمی شوند. اما شبکه های عصبی بازگشتی (RNN) یک لایه بازخورد دارند که در آن خروجی شبکه به همراه ورودی بعدی، به شبکه بازگرداننده می شود. RNN می تواند به علت داشتن حافظه داخلی، ورودی قبلی خود را به خاطر بسپارد و از این حافظه برای پردازش دنباله ایی از ورودی ها استفاده کند. به بیان ساده، شبکه‌های عصبی بازگشتی شامل یک حلقه بازگشتی هستند که موجب می شود اطلاعاتی را که از لحظات قبلی بدست آورده ایم از بین نروند و در شبکه باقی بمانند.



An unrolled recurrent neural network.

شکل ۳: یک بلوک شبکه عصبی بازگشتی

منبع: یافته های پژوهشگر

همانطور که در شکل فوق مشاهده می شود، X_0 را از دنباله ورودی می گیرد، خروجی h_0 به همراه X_1، ورودی مرحله بعدی است. در مرحله بعد، خروجی h_1 و X_2 ورودی مرحله بعد است. به این ترتیب، شبکه در هنگام آموزش قادر به یادآوری ورودی های قبلی خواهد بود.

شبکه عصبی بازگشتی می تواند کاربردهایی به شرح زیر داشته باشد:

۱. شرح نویسی عکس[4]: شبکه عصبی بازگشتی با تحلیل حالت کنونی عکس، برای شرح نویسی عکس به کار می رود.

1. [1] Recurrent Neural Networks
[2] Sequential data
[3] Feed Forward
[4] Image Captioning

۲. پیش بینی سری های زمانی[1]: هر مسئله سری زمانی مانند پیش بینی قیمت یک سهام در یک ماه خاص، با RNN قابل انجام است.

۳. پردازش زبان طبیعی[2]: کاوش متن و تحلیل احساسات می تواند با استفاده از RNN انجام شود.

۴. ترجمه ماشینی[3]: شبکه RNN می تواند ورودی خود را از یک زبان دریافت و آن را به عنوان خروجی به زبان دیگری ترجمه کند.

۵. یافته های پژوهش

آمار توصیفی متغیرهای پژوهش

در جدول ۳، برخی از مفاهیم آمار توصیفی متغیرها، شامل میانگین، میانه،حداقل مشاهدات، حداکثر مشاهدات و انحراف معیار ارائه شده است. با توجه به پنل الف به عنوان مثال نتایج نشان می دهد که، میانگین متغیر حاشیه سود ناخالص ۰.۱۷۹ می باشد که با توجه به انحراف معیار (۰.۱۹۶) از نوسان پذیری بالایی برخوردار است. میانگین متغیر بازده دارایی ها ۰.۱۳۱ می باشد که با توجه به انحراف معیار (۰.۱۳۴) از نوسان پذیری بالایی برخوردار است. میانگین نسبت کل بدهی ها به کل دارایی ها ۰.۵۷۲ می باشد که نشان دهنده این است که به طور متوسط ۵۷.۲٪ منابع مالی شرکت ها از طریق بدهی تأمین مالی شده است. با توجه به پنل ب نیز نتایج نشان می دهد که درصد فراوانی صورت های مالی متقلبانه ۲۹.۸٪ می باشد که نشان دهنده این است که صورت های مالی ۲۹.۸٪ شرکت های موجود در نمونه آماری پژوهش متقلبانه می باشد.

جدول ۳: آمار توصیفی متغیرهای پژوهش

پنل الف: متغیرهای پیوسته							
متغیرها	نماد	تعداد مشاهدات	میانگین	میانه	حداکثر	حداقل	انحراف معیار
حاشیه سود ناخالص	GPM	1800	0.179	0.136	0.616	-0.151	0.196
حاشیه سود خالص	NPM	1800	0.164	0.128	0.566	-0.154	0.183
حاشیه سود عملیاتی	OPM	1800	0.191	0.157	0.528	-0.107	0.172
بازده دارایی ها	ROA	1800	0.131	0.108	0.414	-0.089	0.134
بازده حقوق صاحبان سهام	ROE	1800	0.294	0.291	0.749	-0.236	0.260
نسبت سود قبل از بهره و مالیات	EBIT/A	1800	0.187	0.161	0.466	-0.029	0.139
گردش کل دارایی ها	ASTRN	1800	0.919	0.795	2.067	0.297	0.478
گردش دارایی های ثابت	FXASTRN	1800	6.396	4.479	21.330	0.703	5.679

[1] Time Series Prediction
[2] Natural Language Processing
[3] Machine Translation

پنل الف: متغیرهای پیوسته							
متغیرها	نماد	تعداد مشاهدات	میانگین	میانه	حداکثر	حداقل	انحراف معیار
گردش موجودی کالا	INVTRN	1800	4.008	2.851	12.172	1.007	2.988
گردش حساب های دریافتنی	ACRECTRN	1800	5.416	3.545	20.900	0.925	5.168
نسبت کل بدهی ها به کل دارایی ها	LIB/ASST	1800	0.572	0.579	0.923	0.208	0.200
نسبت کل بدهی ها به حقوق صاحبان سهام	LIB/EQT	1800	1.849	1.268	6.786	0.172	1.701
نسبت سود انباشته به کل دارایی ها	ACUM/ASST	1800	0.148	0.152	0.495	-0.249	0.193
نسبت سود انباشته به سرمایه	ACUM/EQT	1800	0.332	0.401	0.880	-0.723	0.413
نسبت پوشش هزینه بهره	INTCOV	1800	19.244	4.459	153.122	-0.401	37.205
نسبت جاری	CURRAT	1800	1.526	1.348	3.440	0.599	0.717
نسبت آنی	QUICRAT	1800	0.816	0.728	2.067	0.193	0.484
نسبت وجه نقد آزاد به درآمد	FCF/REV	1800	0.054	0.043	0.323	-0.261	0.139
نسبت مخارج سرمایه ای به درآمد	CE/REV	1800	0.045	0.012	0.224	-0.019	0.075
نسبت مخارج سرمایه ای به سود خالص	CE/NP	1800	0.248	0.070	1.652	-0.519	0.600
نسبت وجه نقد عملیاتی به درآمد	OCF/REV	1800	0.136	0.112	0.473	-0.117	0.153
نسبت وجه نقد عملیاتی به بدهی	OCF/LIB	1800	0.247	0.162	1.145	-0.144	0.316
ارزش شرکت	EV	1800	15.635	15.598	18.757	12.944	1.669
ارزش دفتری شرکت	BV	1800	14.695	14.505	17.970	12.255	1.516

پنل ب: متغیرهای دو ارزشی				
متغیرها	نماد	طبقه	فراوانی	درصد فراوانی
صورت های مالی متقلبانه	FSF	1	536	29.8%
		0	1264	70.2%

منبع: یافته های پژوهشگر

جدول ۴: آزمون t در شرکت های دارای صورت های مالی متقلبانه و بدون صورت های مالی متقلبانه

متغیرها	نماد	صورت های مالی متقلبانه (تعداد مشاهدات = ۵۳۶)	بدون صورت های مالی متقلبانه (تعداد مشاهدات = ۱۲۶۴)	آماره t
حاشیه سود ناخالص	GPM	*0.0592*	*0.2292*	-18.315***
حاشیه سود خالص	NPM	*0.0534*	*0.2114*	-17.911***
حاشیه سود عملیاتی	OPM	*0.0739*	*0.2403*	-22.228***

متغیرها	نماد	صورت های مالی متقلبانه (تعداد مشاهدات = ۵۳۶)	بدون صورت های مالی متقلبانه (تعداد مشاهدات = ۱۲۶۴)	آماره t
بازده دارایی ها	ROA	*0.0199*	*0.1779*	-31.566****
بازده حقوق صاحبان سهام	ROE	*0.1082*	*0.3729*	-22.301****
نسبت سود قبل از بهره و مالیات	EBIT/A	*0.0701*	*0.2366*	-32.817***
گردش کل دارایی ها	ASTRN	*0.5773*	*1.0637*	-27.466***
گردش دارایی های ثابت	FXASTRN	*3.5672*	*7.5956*	-16.780****
گردش موجودی کالا	INVTRN	*3.622*	*4.1721*	-3.644***
گردش حساب های دریافتنی	ACRECTRN	*3.5139*	*6.2228*	-11.929***
نسبت کل بدهی ها به کل دارایی ها	LIB/ASST	*0.6622*	*0.535*	12.732***
نسبت کل بدهی ها به حقوق صاحبان سهام	LIB/EQT	*2.312*	*1.653*	6.873***
نسبت سود انباشته به کل دارایی ها	ACUM/ASST	*-0.0111*	*0.2168*	-27.260***
نسبت سود انباشته به سرمایه	ACUM/EQT	*0.0941*	*0.433*	-15.039***
نسبت پوشش هزینه بهره	INTCOV	*5.4553*	*25.0905*	-14.125***
نسبت جاری	CURRAT	*1.1201*	*1.6985*	-19.202***
نسبت آنی	QUICRAT	*0.6192*	*0.8993*	-13.255***
نسبت وجه نقد آزاد به درآمد	FCF/REV	*-0.056*	*0.1007*	-23.443***
نسبت مخارج سرمایه ای به درآمد	CE/REV	*0.06*	*0.0381*	4.967***
نسبت مخارج سرمایه ای به سود خالص	CE/NP	*0.3256*	*0.2151*	3.028****
نسبت وجه نقد عملیاتی به درآمد	OCF/REV	*0.1137*	*0.1461*	-3.929***
نسبت وجه نقد عملیاتی به بدهی	OCF/LIB	*0.0893*	*0.3146*	-19.555***
ارزش شرکت	EV	*15.3186*	*15.7694*	-5.280***
ارزش دفتری شرکت	BV	*14.5811*	*14.744*	-1.996**

* معناداری در سطح اطمینان ۹۰٪، ** معناداری در سطح اطمینان ۹۵٪ و *** معناداری در سطح اطمینان ۹۹٪

منبع: یافته های پژوهشگر

با توجه به جدول ۴ نتایج حاصل از آزمون t نشان می دهد که در سطح اطمینان ۹۵٪ کلیه متغیرهای مستقل (پیش بین) در شرکت های دارای صورت های مالی متقلبانه و بدون صورت های مالی متقلبانه تفاوت معناداری با هم دارند و میانگین متغیرهای حاشیه سود ناخالص، حاشیه سود خالص، حاشیه سود عملیاتی، بازده دارایی ها، بازده

حقوق صاحبان سهام، نسبت سود قبل از بهره و مالیات، گردش کل دارایی ها، گردش دارایی های ثابت، گردش موجودی کالا، گردش حساب های دریافتنی، نسبت سود انباشته به کل دارایی ها، نسبت سود انباشته به سرمایه، نسبت پوشش هزینه بهره، نسبت جاری، نسبت آنی، نسبت وجه نقد آزاد به درآمد، نسبت وجه نقد عملیاتی به درآمد، نسبت وجه نقد عملیاتی به بدهی، ارزش شرکت و ارزش دفتری شرکت در شرکت های دارای صورت های مالی متقلبانه در مقایسه با شرکت های دارای بدون صورت های مالی متقلبانه کمتر است در حالی که، میانگین متغیرهای نسبت کل بدهی ها به کل دارایی ها، نسبت کل بدهی ها به حقوق صاحبان سهام، نسبت مخارج سرمایه ای به درآمد و نسبت مخارج سرمایه ای به سود خالص در شرکت های دارای صورت های مالی متقلبانه در مقایسه با شرکت های دارای بدون صورت های مالی متقلبانه بیشتر است.

۵-۲. فرآیند ایجاد مدل های پژوهش

به منظور پاسخگویی به سوالات و همچنین دستیابی به اهداف پژوهش ابتدا مدل های مورد نظر در تحقیق حاضر را شامل مدل های ایجادی با الگوریتم های حافظه طولانی کوتاه مدت (LSTM)، ماشین بردار پشتیبان(SVM)، شبکه عصبی پیچشی(CNN) و شبکه عصبی بازگشتی (RNN) و با استفاده از متغیرهای انتخاب شده بر اساس آزمون مقایسه میانگین دو نمونه را ایجاد و به مقایسه نتایج ایجاد شده خواهیم پرداخت.

۵-۲-۱. متغیرهای نهایی پژوهش

به منظور تعیین متغیرهای نهایی (با اهمیت) پژوهش از میان متغیرهای اولیه پژوهش (۲۴ متغیر انتخاب شده براساس مبانی نظری و مطالعات تجربی)، از آزمون مقایسه میانگین دو نمونه مستقل استفاده می کنیم. به منظور بررسی مقایسه میانگین دو نمونه، ابتدا باید واریانس دو نمونه را مقایسه نمود به عبارت دیگر آزمون تساوی واریانس ها مقدم بر آزمون تساوی میانگین ها است. جهت تساوی واریانس ها از آزمون لوین که مبتنی بر آماره فیشر است، استفاده می کنیم. در این آزمون نیازی نیست که توزیع داده ها نرمال باشد.

آماره t جهت آزمون تساوی میانگین دو نمونه، در حالت تساوی و عدم تساوی واریانس دو نمونه مورد نظر، محاسبه می شود. در حالت تساوی واریانس ها از رابطه (۲) برای محاسبه آماره t استفاده می کنیم که در این حالت درجه آزادی برابر $df=n_1+n_2-2$ است.

رابطه (۲):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (u_1 - u_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_1}}}$$

رابطه (۳):

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

در حالت عدم تساوی واریانس ها آماره t از رابطه (۴) و درجه آزادی از رابطه (۵) محاسبه می شود.

رابطه (۴):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (u_1 - u_2)}{\sqrt{\frac{S_2^2}{n_1} + \frac{S_2^2}{n_1}}}$$

رابطه (۵):

$$t = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{(\frac{S_1^2}{n_1})^2}{n_1 - 1} + \frac{(\frac{S_2^2}{n_2})^2}{n_2 - 1}}$$

که در آنها n_1 و n_2 تعداد اعضای نمونه اول و نمونه دوم و S_1 و S_2 انحراف معیار نمونه اول و دوم هستند.

سپس براساس عدد t محاسبه شده طبق روابط فوق و عدد t طبق جدول با توجه به سطح خطای ۵ درصد به تفسیر نتایج خواهیم پرداخت، اگر عدد t محاسبه شده از عدد t طبق جدول کوچک تر باشد ($p\text{-value}<0.05$) ، معنادار بودن تفاوت میانگین های دو گروه پذیرفته می شود و در غیر این صورت رد خواهد شد.

با توجه به جدول ۴ نتایج حاصل از آزمون t نشان می دهد که در سطح اطمینان ۹۵٪ کلیه متغیرهای مستقل (پیش بین) در شرکت های دارای صورت های مالی متقلبانه و بدون صورت های مالی متقلبانه تفاوت معناداری با هم دارند. بنابراین در ادامه به منظور پیاده سازی الگوریتم یادگیری عمیق از ۲۴ متغیر پیش بین انتخاب شده براساس آزمون t استفاده می گردد.

۵-۲-۲. اجرای مدل ها

به منظور پیاده سازی و ارزیابی الگوریتم ها باید داده های پژوهش را به دو دسته داده های آموزش و داده های آزمایش تقسیم نماییم. از مجموعه داده های آموزش برای ساخت و از داده های آزمایش نیز برای ارزیابی مدل ایجاد شده در مرحله آموزش بهره خواهیم برد. تکنیک‌های دسته بندی را می‌توان بر اساس معیارهایی مانند صحت، سرعت و پایداری با هم مقایسه نمود. صحت یک روش دسته بندی، بستگی به تعداد پیش بینی‌های درستی که آن مدل انجام می‌دهد دارد. سرعت یک روش دسته بندی زمان لازم برای ساخت و استفاده از مدل در دسته بندی است و پایداری توانایی برخورد مدل در مواجهه با داده های غیر معمول و یا مقادیر مفقود شده را نشان می‌دهد.

به منظور ارزیابی صحت و پایداری مدل‌ها، داده های پژوهش را ۵۰ مرتبه به صورت تصادفی به دو دسته‌ی داده های آموزش و آزمایش تقسیم نموده و به ایجاد مدل و ارزیابی نتایج حاصل از آنها پرداخته‌ایم. به این صورت که در هر مرتبه ۷۵ درصد داده‌ها را برای آموزش مدل و ۲۵ درصد آن را برای ارزیابی و آزمایش مدل مورد استفاده قرار گرفته است و میانگین نتایج حاصل از ۵۰ مرتبه اجرای هر مدل به عنوان نتیجه نهایی آن در نظر گرفته شده است.

جدول ۵: پارامترهای شبیه‌سازی

پارامتر شبیه‌سازی	مقدار
تعداد نمونه‌ها کل دیتاست	۱۸۰۰ سال-شرکت (مشاهده)
تعداد ویژگی هر نمونه	۲۴ ویژگی
تعداد نمونه های آموزش مدل	۱۳۵۰
تعداد نمونه های تست مدل	۴۵۰

منبع: یافته های پژوهشگر

۵-۲-۳. پارامترهای ارزیابی مدل

تشخیص یک روش دسته‌بندی بر روی داده‌ها که تنها دارای دو کلاس می‌باشد با یکی از موارد مثبت درست (TP)، منفی نادرست (FN)، منفی درست (TN) و مثبت نادرست (FP) بیان می‌شود. بر این اساس می‌توان پارامتر صحت(یا دقت)[1] را تعریف نمود. دقت، استانداردترین متریک برای خلاصه سازی عملکرد تشخیص در تمامی‌کلاس‌ها می‌باشد که بصورت فرمول زیر محاسبه می‌شود.

رابطه (۹):

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

همانگونه که در جدول (۶) نشان داده شده است، برای محاسبه دقت دسته‌بندی ابتدا باید ماتریس اغتشاش[2] محاسبه شود. داده‌های روی قطر اصلی، تشخیص‌های درست روش دسته بندی استفاده شده است.

جدول ۶: ماتریس اغتشاش

True Positives (TP)	False Positives (FP)
False Negatives (FN)	True Negatives (TN)

منبع: یافته های پژوهشگر

با توجه به هدف پژوهش حاضر که کشف صورت های مالی متقلبانه است پارامترهای ماتریس اغتشاش به صورت زیر تعریف می شود:

TP: درصد تشخیص درست شرکت های دارای صورت های مالی متقلبانه

[1]- Accuracy
[2]- Confusion Matrix

TN: درصد تشخیص درست شرکت های بدون صورت های مالی متقلبانه

FP: درصد تشخیص غلط شرکت های دارای صورت های مالی متقلبانه به عنوان شرکت های بدون صورت های مالی متقلبانه

FN: درصد تشخیص غلط شرکت های بدون صورت های مالی متقلبانه به عنوان شرکت های دارای صورت های مالی متقلبانه

در ادامه ماتریس اغتشاش هر روش دسته بندی در کشف صورت های مالی متقلبانه نشان داده شده است.

نتایج حاصل از الگوریتم حافظه طولانی کوتاه مدت (LTSM) که در قالب ماتریس اغتشاش و در جدول (۷) ارائه شده است، نشان می دهد که میزان خطای مدل در شناسایی شرکت های دارای صورت های مالی متقلبانه و شرکت های بدون صورت های مالی متقلبانه به ترتیب ۰.۶٪ و ۲.۲٪ می باشد. در حالت کلی نتایج حاصل از الگوریتم LSTM نشان می دهد که میزان دقت این مدل در کشف صورت های مالی متقلبانه شرکت های پذیرفته شده در بورس اراق بهادار تهران تقریباً ۹۸.۶٪ است، این موضوع نشان دهنده این است که با توجه به متغیرهای انتخاب شده به روش T-test و الگوریتم حافظه طولانی کوتاه مدت (LTSM)، می توان با احتمال ۹۸.۶٪ شرکت های دارای صورت های مالی متقلبانه را شناسایی کرد.

جدول ۷: نتایج ماتریس اغتشاش در کشف صورت های مالی متقلبانه (الگوریتم حافظه طولانی کوتاه مدت)

LSTM & T-test		
	Positive	Negative
Positive	0.994	0.022
Negative	0.006	0.978
Accuracy	0.986	

منبع: یافته های پژوهشگر

نتایج حاصل از الگوریتم ماشین بردار پیشتیبان (SVM) که در قالب ماتریس اغتشاش و در جدول (۸) ارائه شده است، نشان می دهد که میزان خطای مدل در شناسایی شرکت های دارای صورت های مالی متقلبانه و شرکت های بدون صورت های مالی متقلبانه به ترتیب 10.9٪ و 12.4٪ می باشد. در حالت کلی نتایج حاصل از الگوریتم SVM نشان می دهد که میزان دقت این مدل در کشف صورت های مالی متقلبانه شرکت های پذیرفته شده در بورس اراق بهادار تهران تقریباً 88.4٪ است، این موضوع نشان دهنده این است که با توجه به متغیرهای انتخاب شده به روش T-test و الگوریتم ماشین بردار پیشتبان (SVM)، می توان با احتمال 88.4٪ شرکت های دارای صورت های مالی متقلبانه را شناسایی کرد.

نتایج حاصل از الگوریتم شبکه عصبی پیچشی (CNN) که در قالب ماتریس اغتشاش و در جدول (۹) ارائه شده است، نشان می دهد که میزان خطای مدل در شناسایی شرکت های دارای صورت های مالی متقلبانه و شرکت های بدون صورت های مالی متقلبانه به ترتیب ۱۸.۳٪ و ۱۹.۱٪ می باشد. در حالت کلی نتایج حاصل از الگوریتم CNN نشان می دهد که میزان دقت این مدل در کشف صورت های مالی متقلبانه شرکت های پذیرفته

شده در بورس اراق بهادار تهران تقریباً ۸۱.۳٪ است، این موضوع نشان دهنده این است که با توجه به متغیرهای انتخاب شده به روش T-test و الگوریتم شبکه عصبی پیچشی (CNN)، می توان با احتمال ۸۱.۳٪ شرکت های دارای صورت های مالی متقلبانه را شناسایی کرد.

جدول ۸: نتایج ماتریس اغتشاش در کشف صورت های مالی متقلبانه (الگوریتم ماشین بردار پشتیبان)

SVM & T-test		
	Positive	Negative
Positive	0.891	0.124
Negative	0.109	0.876
Accuracy	0.884	

منبع: یافته های پژوهشگر

جدول ۹: نتایج ماتریس اغتشاش در کشف صورت های مالی متقلبانه (الگوریتم شبکه عصبی پیچشی)

CNN & T-test		
	Positive	Negative
Positive	0.817	0.191
Negative	0.183	0.809
Accuracy	0.813	

منبع: یافته های پژوهشگر

نتایج حاصل از الگوریتم شبکه عصبی بازگشتی (RNN) که در قالب ماتریس اغتشاش و در جدول (۱۰) ارائه شده است، نشان می دهد که میزان خطای مدل در شناسایی شرکت های دارای صورت های مالی متقلبانه و شرکت های بدون صورت های مالی متقلبانه به ترتیب 5.2٪ و 6.1٪ می باشد. در حالت کلی نتایج حاصل از الگوریتم RNN نشان می دهد که میزان دقت این مدل در کشف صورت های مالی متقلبانه شرکت های پذیرفته شده در بورس اراق بهادار تهران تقریباً ۹۴.۴٪ است، این موضوع نشان دهنده این است که با توجه به متغیرهای انتخاب شده به روش T-test و الگوریتم شبکه عصبی بازگشتی (RNN)، می توان با احتمال ۹۴.۴٪ شرکت های دارای صورت های مالی متقلبانه را شناسایی کرد.

جدول ۱۰: نتایج ماتریس اغتشاش در کشف صورت های مالی متقلبانه (الگوریتم شبکه عصبی بازگشتی)

RNN & T-test		
	Positive	Negative
Positive	0.948	0.061
Negative	0.052	0.939
Accuracy	0.944	

منبع: یافته های پژوهشگر

جدول ۱۱: خلاصه نتایج الگوریتم LSTM و سایر الگوریتم های یادگیری عمیق در کشف صورت های مالی متقلبانه

الگوریتم	Accuracy	FN
LSTM	*0.986*	*0.006*
SVM	*0.884*	*0.109*
CNN	*0.813*	*0.183*
RNN	*0.944*	*0.052*

منبع: یافته های پژوهشگر

با توجه به جدول *۱۱* به طور خلاصه نتایج حاصل از الگوریتم های یادگیری عمیق نشان می دهد که صحت کلی الگوریتم های LSTM، SVM، CNN و RNN به ترتیب 98.6٪، 88.4٪، 81.3٪ و 94.4٪ می باشد که نشان دهنده این است که الگوریتم LSTM بهترین عملکرد و الگوریتم CNN بدترین عملکرد را در کشف شرکت های دارای صورت های مالی متقلبانه دارد. به عبارت دیگر، نتایج نشان دهنده کارآ بودن الگوریتم حافظه طولانی کوتاه مدت (LSTM) نسبت به سایر الگوریتم های یادگیری عمیق می باشد. بنابراین در شرکت های پذیرفته شده در بورس اوراق بهادار تهران الگوریتم LSTM کاراترین مدل را برای کشف صورت های مالی متقلبانه فراهم می کند.

۶. بحث و نتیجه گیری

در سال‌های اخیر، تعدادی از کارشناسان و محققین سعی کرده‌اند از روش‌های یادگیری ماشین و داده‌کاوی برای انجام تحقیقات در این زمینه به عنوان راهی برای کاهش خطاهای تشخیص استفاده کنند. در حالی که بسیاری از استراتژی های تجاری مبتنی بر دقت صورت های مالی هستند، منابع کافی برای تجزیه و تحلیل همه آنها به طور کامل وجود ندارد. از آنجایی که حسابرسان در چندین نمونه از صورت های مالی متقلبانه مقصر شناخته شده‌اند، مدل‌های کشف تقلب مالی معتبر باید به روشی آسان برای حسابرسان، سرمایه‌گذاران، سیاست گذاران و سایر ذینفعان ارائه شوند. به دلیل اتکای ذاتی به مفروضات توزیعی محدود، تکنیک‌های پارامتریک مانند رگرسیون لجستیک (LR) فاقد کاربرد کلی است که رویکردهای غیر پارامتری ممکن است ارائه کنند (اسچرییبر-گریگوری و بادر[1]، ۲۰۱۸). از سوی دیگر، تحقیقات موجود در مورد روش‌های کمی کشف تقلب مالی، عمدتاً بر صنعت بانکداری و خدمات مالی متمرکز شده است، که عمدتاً بر روی کشف تقلب بیمه و کارت اعتباری است. تکنیک های کمی برای شناسایی و عدم تمایل به صورت های مالی متقلبانه باید در مجلات معتبر دانشگاهی مورد توجه قرار گیرد. ادبیات علمی و آکادمیک فعلی در حال حاضر به دنبال قوانین یا طبقه بندی های بیشتر از داده های قبلی برای دستیابی به هدف پیش بینی یا تشخیص است. یادگیری ماشینی با مقدار مناسب داده می‌تواند نتایج دقیق‌تری را در مقایسه با رویکرد سنتی پیش‌بینی و طبقه‌بندی کند.

کشف تقلب در صورت‌های مالی در چند دهه گذشته با تاکید بر ناهنجاری‌های حسابداری به طور کلی و به طور خاص بر تقلب در صورت‌های مالی مورد توجه قرار گرفته است. در حالی که تحقیقات اولیه از تکنیک‌های

[1] Schreiber-Gregory and Bader

آماری یا سنتی استفاده می‌کرد که هم زمان‌بر و هم پرهزینه هستند، اخیراً با ظهور داده‌های بزرگ و یادگیری ماشین تمرکز بیشتر شده است (محمدی و همکاران[۱]، ۲۰۲۰). رویکردهای آماری بر روش‌های ریاضی سنتی متمرکز هستند، در حالی که روش‌های مربوط به یادگیری ماشین بر یادگیری و هوش متمرکز هستند. در حالی که هر دو دسته شباهت‌های زیادی دارند، تفاوت اصلی بین آنها این است که روش‌های آماری سخت‌تر و غیر منعطف هستند، در حالی که روش‌های یادگیری ماشین قادر به یادگیری و تطبیق با حوزه مسأله هستند (هامفریس و همکاران[۲]، ۲۰۱۱).

علاوه بر این، مطالعات قبلی کارایی برتر رویکردهای یادگیری ماشین را نسبت به رویکردهای آماری مرسوم نشان داده‌اند (وست و همکاران[۳]، ۲۰۱۴). با توجه به ادبیات، هیچ استراتژی واحدی برای شناسایی تقلب در صورت‌های مالی وجود ندارد (هامال و سنوار[۴]، ۲۰۲۱). یافته‌های پیشین (کراجا و همکاران[۵]، ۲۰۲۰) نشان می‌دهد که مدل‌هایی که با استفاده از رویکردهای یادگیری ماشین ساخته شده‌اند، می‌توانند به طور مؤثر تقلب در صورت‌های مالی را شناسایی کنند، با تکامل مستمر رفتار متقلبانه گزارشگری مالی همگام باشند و با به‌روزترین فناوری پاسخ دهند. در پژوهشی گوپتا و مهتا[۶] (۲۰۲۱)، نشان دادند که مدل‌های تشخیص توسعه‌یافته با استفاده از رویکردهای یادگیری ماشین دقیق‌تر از روش‌های سنتی هستند. بنابراین، در پژوهش حاضر بررسی ما از تحقیقات گذشته محدود به مقالاتی است که تنها از تکنیک‌های یادگیری ماشین برای تشخیص صورت های مالی متقلبانه استفاده کرده‌اند. بسیاری از تحقیقات در ادبیات قبلی تشخیص صورت های مالی متقلبانه را به عنوان یک مسئله طبقه بندی باینری، گاهی به عنوان یک مسئله چند کلاسه و گاهی اوقات به عنوان یک مشکل خوشه بندی فرموله کرده اند. محققان هر دو تجزیه و تحلیل کمی و کیفی صورت های مالی متقلبانه را انجام داده اند. متن کاوی به طور گسترده برای تحقیقات کیفی استفاده شده است. در پژوهش حاضر، ما بر روی مقالاتی تمرکز می کنیم که تجزیه و تحلیل کمی را با استفاده از تکنیک های یادگیری ماشین انجام می دهند. در مراحل اولیه، تحقیقات عمدتاً شامل رویکردهای شبکه عصبی(NN) (سچینی و همکاران[۷]، ۲۰۱۰؛ پای و همکاران[۸]، ۲۰۱۱؛ الفیض و فاتی[۹]، ۲۰۲۲؛ استرلسنیا و پراکونویت[۱۰]، ۲۰۲۳؛ کومار و همکاران[۱۱]، ۲۰۲۲)، رگرسیون لجستیک(LR)، درخت تصمیم(DT)، ماشین بردار پیشتیبان (SVM)، تحلیل متمایز(DA) و شبکه باور بیزی[۱۲] بود. تکنیک‌های یادگیری نظارت شده برای تجزیه و تحلیل بیشتر از روش‌های بدون نظارت انتخاب شده‌اند، با مطالعاتی که از ایالات متحده آمریکا، چین،

[1] Mohammadi et al.
[2] Humpherys et al.
[3] West et al.
[4] Hamaland Senvar
[5] Craja et al.
[6] Gupta and Mehta
[7] Cecchini et al.
[8] Pai et al.
[9] Alfaiz and Fati
[10] Strelcenia and Prakoonwit,
[11] Kumar et al.
[12] Bayesian Belief Network

تایوان و اسپانیا انجام شده است، ۶۵ درصد از این مقالات را تشکیل می‌دهند (البا‌شراوی[1]، ۲۰۱۶). تعداد قابل توجهی از مطالعات عملکرد طبقه‌بندی‌کننده‌ها را در تشخیص صورت های مالی متقلبانه تجزیه و تحلیل کرده‌اند و نشان می‌دهند که ماشین بردار پشتبان(سچینی و همکاران، ۲۰۱۰؛ پای و همکاران، ۲۰۱۱؛ پرولس[2]، ۲۰۱۱؛ آسیمیت و همکاران[3]، ۲۰۲۲؛ موئیپیا و همکاران[4]، ۲۰۱۴؛ یائو و همکاران[5]، ۲۰۱۹)، شبکه عصبی(هان[6]، ۲۰۱۷؛ لین و همکاران[7]، ۲۰۱۵؛ راویسانکار و همکاران[8]، ۲۰۱۱؛ ریزکی و همکاران[9]، ۲۱۷؛ مورورونخر و همکاران[10]، ۲۰۲۲؛ پیرز لوپز و همکاران[11]، ۲۰۱۹) و درخت تصمیم(گوپتا و گیل[12]، ۲۰۱۲؛ چن[13]، ۲۰۱۶) در کشف/پیش‌بینی صورت های مالی متقلبانه عملکرد خوبی دارند.

با ظهور یادگیری عمیق، مدل های جدید برای تحلیل و پیش بینی سریهای زمانی توسعه یافته اند. شبکه های عصبی مکرر(روملهارت و همکاران[14]، ۱۹۸۶) بیشتر، کارآمدترین روش پیش بینی سری زمانی اند(لیو و همکاران[15]، ۲۰۲۰). درواقع، RNN، شبکه مصنوعی است که در آن گره ها به صورت حلقه وصل می شوند و حالت داخلی شبکه می تواند رفتار زمان بندی پویا را به نمایش بگذارد. مهم ترین ضعف شبکه های مبتنی بر RNN در هنگام یادگیری وابستگی های طولانی مدت است. برای غلبه بر این مشکل، LSTM و واحدهای مکرر گیت دار[16] ارائه شدند که عملیات خطی ساده را روی اطلاعات نورون انجام می دهند و اطلاعات خارجی از لحظه فعلی را ازطریق مکانیزم گیت اضافه می کنند(چونگ و همکاران[17]، ۲۰۱۴). با وجود مزیت های شبکه های LSTM، هنوز عملکرد آنها بر داده های پیش بینی سری زمانی، رضایت بخش نیست. معماری های کم عمق ویژگی های داده های سری زمانی را به طور کارآمد نشان نمی دهند؛ به خصوص زمانی که داده های سری زمانی با فواصل طولانی مدت و بسیار غیرخطی پردازش می شوند (سغیر و کتب[18]، ۲۰۱۹). همچنین با پیچیده تر شدن معماری شبکه های عمیق، یک سؤال پیش می آید که چطور یک شبکه را تنظیم کنیم؛ البته می توان تعداد محدودی از ابرپارامترها را با آزمایش بهینه کارد ؛ اما شبکه های عمیق دارای توپولوژی پیچیده و صدها ابرپارامترند. اغلب موفقیت در حل مسئله به

[1] Albashrawi
[2] Perols
[3] Asimit et al.
[4] Moepya et al.
[5] Yao et al.
[6] Han
[7] Lin et al.
[8] Ravisankaret al.
[9] Rizki et al.
[10] Murorunkwere et al.
[11] Pérez López et al.
[12] Gupta and Gill
[13] Chen
[14] Rumelhart et al.
[15] Liu et al.
[16] Gated Recurrent Unit
[17] Chung et al.
[18] Sagheer and Kotb

انتخاب معماری مناسب برای آن مسئله بستگی دارد(میکولینن و همکاران[1]، ۲۰۱۹). در همین راستا، پژوهش حاضر به دنبال پاسخی به این سؤال است که چگونه تکنیک یادگیری عمیق حافظه طولانی کوتاه مدت (LSTM) قدرت کشف صورت های مالی متقلبانه را در شرکت های پذیرفته شده در بورس اوراق بهادار تهران را بهبود می بخشد؟ به عبارت دیگر، صحت و دقت تکنیک یادگیری عمیق حافظه طولانی کوتاه مدت (LSTM) در کشف صورت های مالی متقلبانه در شرکت های پذیرفته شده در بورس اوراق بهادار تهران چه میزان می باشد؟

از طریق تحلیل اطلاعات مربوط به ۱۵۰ شرکت پذیرفته شده در بورس اوراق بهادار تهران در طی سال های ۱۳۹۱ تا ۱۴۰۲ نتایج حاصل از الگوریتم های یادگیری عمیق نشان می دهد که صحت کلی الگوریتم های LSTM، SVM، CNN و RNN به ترتیب 98.6٪، 88.4٪، 81.3٪ و 94.4٪ می باشد که نشان دهنده االگوریتم LSTM بهترین عملکرد و الگوریتم CNN بدترین عملکرد را در کشف شرکت های دارای صورت های مالی متقلبانه دارد. به عبارت دیگر، نتایج نشان دهنده کارآ بودن الگوریتم حافظه طولانی کوتاه مدت (LSTM) نسبت به سایر الگوریتم های یادگیری عمیق می باشد. بنابراین در شرکت های پذیرفته شده در بورس اوراق بهادار تهران الگوریتم LSTM کاراترین مدل را برای کشف صورت های مالی متقلبانه فراهم می کند. این نتایج تا حدودی با یافته های بدست آمده در پژوهش های کاظمی و پیری(۱۴۰۱)، بهشتی مسئله گو و همکاران (۱۴۰۱)، سیاهکارزاده (۱۴۰۰)، زهیر و همکاران (۲۰۲۳)، علی و همکاران (۲۰۲۳) همخوانی دارد.

این پژوهش می تواند چندین دستاورد علمی به شرح زیر می تواند داشته باشد. اولاً، از دیدگاه نظری، با توجه به بررسی های صورت گرفته از پژوهش های تجربی صورت گرفته در ایران، این پژوهش اولین موردی است که با استفاده از الگوریتم حافظه طولانی کوتاه مدت به کشف صورت های مالی متقلبانه در شرکت های پذیرفته شده در بورس اوراق بهادار تهران پرداخته است. دوماً، با توجه به اینکه کیفیت و درستی صورت های مالی شرکت یکی از متغیرهای کلیدی در تصمیمات سرمایه گذاران است این پژوهش با کشف صورت های مالی متقلبانه با بکارگیری الگوریتم حافظه طولانی کوتاه مدت و سایر الگوریتم های یادگیری عمیق، سهامداران، تحلیلگران و اعتباردهندگان را در اتخاذ تصمیمات اثربخش و کارآمد یاری می کند و در نتیجه باعث افزایش سطح کارآیی بازار در بلندمدت می گردد. سوماً، برای نهادهای نظارتی و تدوین کنندگان قوانین و استانداردها مسئله تقلب به ویژه صورت های مالی متقلبانه همواره در کانون توجه بوده است یافته های این پژوهش می تواند به سازمان حسابرسی، سازمان بورس و اوراق بهادار و سایرهای نهادهای نظارتی رهنمود های لازم را در زمینه تدوین قوانین و استاندارد ارائه دهد. چهارماً، نتایج این پژوهش می تواند حسابرسان را در زمینه گردآوری شواهد و ارزیابی شواهد یاری کند. در نهایت یافته های این پژوهش می تواند ایده های جدیدی در اختیار پژوهشگران و دانشگاهیان قرار دهد.

[1] Miikkulainen et al

فهرست منابع

ابراهیمی، سید بابک، جهانگیرزاده، جواد، کتابیان، حمید. (۱۳۹۵). شناسایی پارامترهای تاثیرگذار بر تقلب در حوزه مالی. فصلنامه مجلس و راهبرد، ۱۲(۸۶). ۱۷۴-۱۴۹.

احمدی، سید جلال، فغانی ماکرانی، خسرو، فاضلی، نقی. (۱۳۹۹). ارائه مدلی برای پیش‌بینی صورت‌های‌مالی متقلبانه و مقایسه صورت‌ها و نسبت‌های مالی با قانون بنفورد. دانش حسابداری و حسابرسی مدیریت، ۹(۳۵)، ۲۲۱-۲۳۷.

اعتمادی، حسین، عبدلی، لیلا. (۱۳۹۶). کیفیت حسابرسی و تقلب در صورت‌های مالی. دانش حسابداری مالی، ۴(۴)، ۲۳-۴۳.

بابانژاد باقری، سیده مریم، پورآقاجان، عباسعلی، عباسیان فریدونی، محمدمهدی. (۱۴۰۲). پیش بینی ارزش شرکت مبتنی بر روش های یادگیری عمیق. فصلنامه اقتصاد مالی، ۱۷(۳)، ۲۹۱-۳۱۸.

برنا، محمدرضا، برادران حسن زاده، رسول، فضل زاده، علیرضا، بادآورنهندی، یونس. (۱۴۰۱). الگوی عوامل موثر بر وقوع تقلب در صورت های مالی با رویکرد حسابداری دادگاهی: بر اساس روش تحلیل مضمون(تم). پژوهش های کاربردی در گزارشگری مالی، ۱۱(۲۱)، ۳۲۰-۲۸۱.

بهشتی مسئله گو، سیده مژگان، افشار کاظمی، محمد علی، حقیقت منفرد، جلال، رضاییان، علی. (۱۴۰۱). یادگیری عمیق برای پیش‌بینی بازار سهام با استفاده از اطلاعات عددی و متنی (رویکرد الگوریتم حافظه کوتاه مدت ماندگار LSTM). مهندسی مالی و مدیریت اوراق بهادار، پذیرفته شده، انتشار آنلاین از تاریخ ۲۵ مرداد ۱۴۰۱.

پسندیده فرد، فایزه، وادی زاده، کاظم، سپاسی سحر. (۱۳۹۹). شناسایی عوامل مؤثر بر گزارشگری مالی متقلبانه و نادرست با استفاده از روش فراترکیب. عنوان نشریه. ۵ (۹)، ۳۰۱-۳۳۴.

پور حیدری، امید، بذر افشان، سعید. (۱۳۹۰). اهمیت بسترهای خطر تقلب از دیدگاه حسابرسان مستقل. پژوهش های حسابداری مالی و حسابرسی، ۳(۱۰)، ۱-۲۶.

خلیلی ثمرین، فاطمه، خلیل پور، مهدی، و رمضانی، جواد. (۱۴۰۰). رتبه بندی عوامل درون سازمانی و برون سازمانی موثر بر گزارشگری مالی متقلبانه با استفاده از فرآیند تحلیل سلسله مراتبی. اقتصاد مالی (اقتصاد مالی و توسعه)، ۱۵(۵۵)، ۲۳۱-۲۴۶.

خواجوی، شکرالله، ابراهیمی، مهرداد. (۱۳۹۶). ارائۀ یک رویکرد محاسباتی نوین برای پیش‌بینی تقلب در صورت‌های مالی، با استفاده از شیوه‌های خوشه‌بندی و طبقه‌بندی (شواهدی از شرکت‌های پذیرفته‌شده در بورس اوراق بهادار تهران). پیشرفت‌های حسابداری، ۹(۲)، ۱-۳۴.

خواجوی، شکراله. (۱۳۹۶). مدل سازی متغیرهای اثرگذار برای کشف تقلب در صورت های مالی با استفاده از تکنیک های داده کاوی. حسابداری مالی، ۹(۳۳)، ۲۳-۵۰.

رحیمیان، نظام الدین، حاجی حیدری، راضیه. (۱۳۹۸). کشف تقلب با استفاده از مدل تعدیل شده بنیش و نسبت‌های مالی. پژوهش های تجربی حسابداری، ۹(۱)، ۴۷-۷۰.

رضائی، مهدی، ناظمی اردکانی، مهدی، ناصر صدرآبادی، علیرضا. (۱۳۹۹). کشف تقلب صورت‌های مالی با توجه به گزارش حسابرسی صورت‌های مالی. حسابداری مدیریت، ۱۳(۴۵)، ۱۴۱-۱۵۳.

رضائی، مهدی، ناظمی اردکانی، مهدی، ناصر صدرآبادی، علیرضا. (۱۴۰۰). پیش بینی تقلب صورت های مالی با استفاده از رویکرد کریسپ (CRISP). دانش حسابداری و حسابرسی مدیریت، ۱۰(۴۰)، ۱۵۰-۱۳۵.

سجادی، سید حسین، کاظمی، توحید. (۱۳۹۵). الگوی جامع گزارشگری مالی متقلبانه در ایران به روش نظریه پردازی زمینه بنیان. پژوهش های تجربی حسابداری، ۶(۳)، ۱۸۵-۲۰۴.

سیاهکارزاده، علی محمد. (۱۴۰۰). پیش‌بینی بازارهای مالی با استفاده از یادگیری عمیق. پایان نامه کارشناسی ارشد، مهندسی کامپیوتر، دانشگاه صنعتی شاهرود.

شکوری، محمدمهدی، طاهرآبادی، علی اصغر، قنبری، مهرداد، جمشیدی، نوید بابک. (۱۴۰۱). تبیین مدل پنتاگون تقلب و ارائه مدل جامع گزارشگری مالی متقلبانه. دانش حسابرسی. ۲۲ (۸۷)،۱۲۰-۱۵۹.

شمس، زهرا، ملانظری، مهناز. (۱۳۹۸). مدل عوامل مؤثر بر قضاوت و تصمیم‌گیری در خصوص تقلب مرتبط با دارایی‌ها. پژوهش های تجربی حسابداری، ۹(۳)، ۲۷۷-۳۰۰.

صفرزاده، محمدحسین. (۱۳۹۱). توانایی نسبتهای مالی در کشف تقلب در گزارشگری مالی: تحلیل لاجیت. مجله دانش حسابداری، ۱(۱)، ۱۳۷-۱۶۳.

صفی خانی، فرهاد، یعقوب نژاد، احمد، جهانشاد، آزیتا. (۱۴۰۲). پیشایندهای گزارشگری مالی متقلبانه در شرکت های دارای بحران مالی. دانش حسابداری و حسابرسی مدیریت، ۱۲(۴۵)، ۲۵۱-۲۷۰.

فرج زاده دهکردی، حسن، آقایی، لیلا. (۱۳۹۴). سیاست تقسیم سود و گزارشگری مالی متقلبانه. مطالعات تجربی حسابداری مالی، ۱۲(۴۵)، ۱۱۴-۹۷.

فرقان دوست حقیقی، کامبیز، و برواری، فرید. (۱۳۸۸). بررسی کاربرد روش های تحلیلی در ارزیابی ریسک تحریف صورت های مالی (تقلب مدیریت). دانش و پژوهش حسابداری، -(۱۶)، ۱-۱۵.

فلاح حمیدی، حامی، فغانی ماکرانی، خسرو، و ذبیحی، علی. (۱۴۰۰). قدرت و بیش اطمینانی مدیران و گزارشگری متقلبانه. اقتصاد مالی (اقتصاد مالی و توسعه)، ۱۵(۵۴)، ۳۷۹-۴۰۰.

کاظمی، توحید، پیری، پرویز. (۱۴۰۱). پیش بینی طرح تقلب در گزارشگری مالی با استفاده از رویکرد یادگیری ماشین در فضای چند کلاسه. پژوهش های تجربی حسابداری، ۱۲(۴۶)، ۲۷۶-۲۵۵.

مرادی، جواد، رستمی، راحله، زارع، رضا. (۱۳۹۳). شناسایی عوامل خطر مؤثر بر احتمال وقوع تقلب در گزارشگری مالی از دید حسابرسان و بررسی تأثیر آن‌ها بر عملکرد مالی شرکت. پیشرفت‌های حسابداری، ۶(۱)، ۱۴۱-۱۷۳.

ملکی کاکلر، حسن، بحری ثالث، جمال، جبارزاده کنگرلویی، سعید، آشتاب، علی. (۱۴۰۰). کارایی مدل‌های آماری والگوهای یادگیری ماشین در پیش بینی گزارشگری مالی متقلبانه. فصلنامه اقتصاد مالی، ۱۵(۵۴)، ۲۶۷-۲۹۲.

ملکی کاکلر، حسن، بحری ثالث، جمال، جبارزاده کنگرلویی، سعید، آشتاب، علی، کاظمی، توحید. (۱۴۰۰). ارائه مدل توسعه یافته ی پنج ضلعی تقلب در گزارشگری مالی با تاکید بر ساختار کنترلهای داخلی. دانش حسابرسی. ۲۱ (۸۴)، ۵۶-۷۹.

مهدوی، غلامحسین، قهرمانی، علیرضا. (۱۳۹۶). ارائه الگویی برای کشف تقلب به وسیله حسابرسان با استفاده از شبکه عصبی مصنوعی. دانش حسابرسی، ۱۷ (۶۷)، ۴۵-۷۰.

میرفندرسکی، سید محمدتقی، قادری، بهمن، آخوندی، مجید. (۱۳۹۷). رابطه ی بین دوره ی تصدی حسابرس و خطر تقلب در گزارشگری مالی با کنترل اثر شرایط نااطمینانی محیطی. فصلنامه حسابدار رسمی، ۲۱(۴۳)، ۵۶-۶۵.

یوخنه القیانی، ماریام، بحری ثالث، جمال، جبارزاده کنگرلوئی، سعید، زواری رضایی، اکبر. (۱۴۰۰). تبیین گزارشگری مالی- مالیاتی متقلبانه شرکت‌ها: رویکرد ترکیبی داده‌کاوی کلاسیک، ANFIS و الگوریتم‌های فراابتکاری. مطالعات تجربی حسابداری مالی، ۱۸(۷۱)، ۸۷-۱۱۲.

یوخنه القیانی، ماریام، جبارزاده کنگرلوئی، سعید، بحری ثالث، جمال، و زواری رضایی، اکبر. (۱۳۹۹). تبیین الگوی تقلب مالیاتی در ایران مبتنی بر رویکرد آمیخته (واکاوی کنش مالیاتی متقلبانه اشخاص حقوقی). دانش حسابرسی، ۲۰(۸۰)، ۲۴۱-۲۸۰.

Albashrawi, M. Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015. J. Data Sci. 2016, 14, 553–569.

Alfaiz, N.S.; Fati, S.M. Enhanced Credit Card Fraud Detection Model Using Machine Learning. Electronics 2022, 11, 662.

Alsinglawi, M.M.A.S.M.A.O.; Almari, M.O.S. Predicting Fraudulent Financial Statements Using Fraud Detection Models. Acad. Strateg. Manag. 2021, 20, 1–17.

Amar, I.A.A.B.; Jarboui, A. Detection of Fraud in Financial Statements: French Companies as a Case Study. Int. J. Acad. Res. Bus. Soc. Sci. 2013, 3, 456–472.

Andrew, C.; Robin. Detecting Fraudulent of Financial Statements Using Fraud S.C.O.R.E Model and Financial Distress. Int. J. Econ. Bus. Account. Res. (IJEBAR) 2022, 6, 211–222.

Asimit, A.V.; Kyriakou, I.; Santoni, S.; Scognamiglio, S.; Zhu, R. Robust Classification via Support Vector Machines. Risks 2022, 10, 154.

Bao, Y.; Ke, B.; Li, B.; Yu, Y.J.; Zhang, J. Detecting accounting fraud in publicly traded US firms using a machine learning approach. J. Account. Res. 2020, 58, 199–235.

Beneish, M.D. The detection of earnings manipulation. Financ. Anal. J. 1999, 55, 24–36.

Cecchini, M.; Aytug, H.; Koehler, G.J.; Pathak, P. Detecting management fraud in public companies. Manag. Sci. 2010, 56, 1146–1160.

Chen, S. Detection of fraudulent financial statements using the hybrid data mining approach. SpringerPlus 2016, 5, 1–16.

Craja, P.; Kim, A.; Lessmann, S. Deep learning for detecting financial statement fraud. Decis. Support Syst. 2020, 139, 113421.

Deebak, B.; Memon, F.H.; Dev, K.; Khowaja, S.A.; Wang, W.; Qureshi, N.M.F. TAB-SAPP: A trust-aware blockchain-based seamless authentication for massive IoT-enabled industrial applications. IEEE Trans. Ind. Inform. 2022, 19, 243–250.

El-Bannany, M.; Sreedharan, M.; Khedr, A.M. A robust deep learning model for financial distress prediction. Int. J. Adv. Comput. Sci. Appl. 2020, 11, 170–175.

Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 2014, 15, 3133–3181.
Gorenc, M. Empirical evidence of financial statement manipulation during economic recessions. Management 2019, 14, 19–31.
Grove, H.; Basilico, E. Fraudulent Financial Reporting Detection Key Ratios Plus Corporate Governance Factors. Int. Stud. Mgt. Org. 2008, 38, 10–42.
Gu, Q.; Zhu, L.; Cai, Z. Evaluation measures of the classification performance of imbalanced data sets. In Proceedings of the Computational Intelligence and Intelligent Systems: 4th International Symposium, ISICA 2009, Huangshi, China, 23–25 October 2009; Proceedings 4. Springer: Cham, Switzerland, 2009; pp. 461–471.
Gupta, R.; Gill, N.S. Prevention and detection of financial statement fraud–An implementation of data mining framework. Editor. Pref. 2012, 3, 150–160
Gupta, S.; Mehta, S.K. Data mining-based financial statement fraud detection: Systematic literature review and meta-analysis to estimate data sample mapping of fraudulent companies against non-fraudulent companies. Glob. Bus. Rev. 2021, 1–26.
Hajek, P.; Henriques, R. Mining corporate annual reports for intelligent detection of financial statement fraud—A comparative study of machine learning methods. Knowl.-Based Syst. 2017, 128, 139–152.
Hamal, S.; Senvar, Ö. Comparing performances and effectiveness of machine learning classifiers in detecting financial accounting fraud for Turkish SMEs. Int. J. Comput. Intell. Syst. 2021, 14, 769–782.
Han, D. Researches of Detection of Fraudulent Financial Statements Based on Data Mining. J. Comput. Theor. Nanosci. 2017, 14, 32–36.
Humpherys, S.L.; Moffitt, K.C.; Burns, M.B.; Burgoon, J.K.; Felix, W.F. Identification of fraudulent financial statements using linguistic credibility analysis. Decis. Support Syst. 2011, 50, 585–594.
Kanapickiene, R.; Grundiene, Z. The Model of Fraud Detection in Financial Statements by Means of Financial Ratios. Procedia Soc. Behav. Sci. 2015, 213, 321–327
Kulikova, L.; Satdarova, D. Internal control and compliance-control as effective methods of management, detection and prevention of financial statement fraud. Acad. Strateg. Manag. J. 2016, 15, 92.
Kumar, R.; Tripathi, R. Secure healthcare framework using blockchain and public key cryptography. In Blockchain Cybersecurity, Trust and Privacy; Springer: Cham, Switzerland, 2020; pp. 185–202.
Kumar, S.; Ahmed, R.; Bharany, S.; Shuaib, M.; Ahmad, T.; Tag Eldin, E.; Rehman, A.U.; Shafiq, M. Exploitation of Machine Learning Algorithms for Detecting Financial Crimes Based on Customers' Behavior. Sustainability 2022, 14, 13875.
Li, H.; Wong, M.L. Financial fraud detection by using Grammar-based multi-objective genetic programming with ensemble learning. In Proceedings of the 2015 IEEE Congress on Evolutionary Computation (CEC), Sendai, Japan, 25–28 May 2015; IEEE: New York, NY, USA, 2015; pp. 1113–1120.
Lin, C.C.; Chiu, A.A.; Huang, S.Y.; Yen, D.C. Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. Knowl.-Based Syst. 2015, 89, 459–470.
Misstatements; Bertomeu, J.; Cheynel, E.; Floyd, E.; Pan, W. Ghost in the Machine: Using Machine Learning to Uncover Hidden; Springer: Cham, Switzrland, 2018; Volume 4, pp. 233–241.
Moepya, S.O.; Akhoury, S.S.; Nelwamondo, F.V. Cost-sensitive classification for financial fraud detection under high classimbalance. In Proceedings of the 2014 IEEE international conference

on data mining workshop, Shenzhen, China, 14–17 December 2014; IEEE: New York, NY, USA, 2014; pp. 183–192.
Mohammadi, M.; Yazdani, S.; Khanmohammadi, M.H.; Maham, K. Financial reporting fraud detection: An analysis of data mining algorithms. Int. J. Financ. Manag. Account. 2020, 4, 1–12.
Murorunkwere, B.F.; Tuyishimire, O.; Haughton, D.; Nzabanita, J. Fraud Detection Using Neural Networks: A Case Study of Income Tax. Future Internet 2022, 14, 168.
Pai, P.F.; Hsu, M.F.; Wang, M.C. A support vector machine-based model for detecting top management fraud. Knowl.-Based Syst. 2011, 24, 314–321.
Paulo Sérgio Gomes Macedo, H.C.I.; Vieira, E.S. A model to detect financial statement fraud in Portuguese companies by the auditor. Contaduría Adm. 2022, 67, 185–209.
Pérez López, C.; Delgado Rodríguez, M.; de Lucas Santos, S. Tax Fraud Detection through Neural Networks: An Application Using a Sample of Personal Income Taxpayers. Future Internet 2019, 11, 86.
Perols, J. Financial statement fraud detection: An analysis of statistical and machine learning algorithms. Audit. J. Pract. Theory 2011, 30, 19–50.
Pintelas, P.; Livieris, I. Ensemble learning and their applications. Algorithms 2020, 1–184.
Ragab, Y. Financial Ratios and Fraudulent Financial Statements Detection: Evidence from Egypt. Int. J. Acad. Res. 2017, 4, 1–6.
Ravisankar, P.; Ravi, V.; Rao, G.R.; Bose, I. Detection of financial statement fraud and feature selection using data mining techniques. Decis. Support Syst. 2011, 50, 491–500.
Rizki, A.A.; Surjandari, I.; Wayasti, R.A. Data mining application to detect financial fraud in Indonesia's public companies. In Proceedings of the 2017 3rd International Conference on Science in Information Technology (ICSITech), Bandung, Indonesia, 25–26 October 2017; IEEE: New York, NY, USA, 2017; pp. 206–211
Saville, A. Using Benford's Law to Detect Data Error and Fraud: An Examination Of Companies Listed on the Johannesburg Stock Exchange. SAJEMS 2006, 9, 341–354
Schreiber-Gregory, D.; Bader, K. Logistic and Linear Regression Assumptions: Violation Recognition and Control. In Proceedings of the SESUG Conference, St. Pete Beach, FL, USA, 14–17 October 2018; pp. 1–6.
Song, X.P.; Hu, Z.H.; Du, J.G.; Sheng, Z.H. Application of machine learning methods to risk assessment of financial statement fraud: Evidence from China. J. Forecast. 2014, 33, 611–626.
Sreedharan, M.; Khedr, A.M.; El Bannany, M. A Multi-Layer Perceptron Approach to Financial Distress Prediction with Genetic Algorithm. Autom. Control. Comput. Sci. 2020, 54, 475–482.
Strelcenia, E.; Prakoonwit, S. Improving Classification Performance in Credit Card Fraud Detection by Using New Data Augmentation. AI 2023, 4, 172–198.
Tilden, C.; Janes, T. Benford's Law as a Useful Tool to Determine Fraud in Financial Statements. J. Financ. Account. 2012, 14, 1–15.
Wadhwa, A.V.K.; Kumar, S. Financial Fraud Prediction Models: A Review of Research Evidence. Int. J. Sci. Technol. Res. 2020, 9, 677–680.
West, J.; Bhattacharya, M.; Islam, R. Intelligent financial fraud detection practices: An investigation. In Proceedings of the International Conference on Security and Privacy in Communication Networks, Beijing, China, 24–26 September 2014; Springer: Cham, Switzrland, 2014; pp. 186–203.
Whiting, D.G.; Hansen, J.V.; McDonald, J.B.; Albrecht, C.; Albrecht, W.S. Machine learning methods for detecting patterns of management fraud. Comput. Intell. 2012, 28, 505–527.
Yao, J.; Pan, Y.; Yang, S.; Chen, Y.; Li, Y. Detecting fraudulent financial statements for the sustainable development of the socio-economy in China: A multi-analytic approach. Sustainability 2019, 11, 1579.

Yao, J.; Zhang, J.; Wang, L. A financial statement fraud detection model based on hybrid data mining methods. In Proceedings of the 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 26–28 May 2018; IEEE: New York, NY, USA, 2018; pp. 57–61.

Zaheer, S.; Anjum, N.; Hussain, S.; Algarni, A.D.; Iqbal, J.; Bourouis, S.; Ullah, S.S. A Multi-Parameter Forecasting for Stock Time Series Data Using LSTM and Deep Learning Model. Mathematics 2023, 11, 590. https://doi.org/10.3390/ math11030590.

Financial Economics
Vol. (19) Issue (72) October 2025

Abstract

Detection of Fraudulent Financial Statements based on Long Short-Term Memory (LSTM) Deep Learning Technique

Javad Sharafkhani[1]
Mohammad Ali Bidari[2]

Received: 20/ July /2025 Accepted: 20/ September /2025

Abstract

The purpose of the current research is to detection of fraudulent financial statements based on long short-term memory deep learning technique in listed companies on the Tehran Stock Exchange and compare the results with other deep learning algorithms. In order to achieve this purpose, 1,800 firm-year (150 firms for 12 years) of observation collected from the annual financial reports of listed companies to the Tehran Stock Exchange during the period 2012 to 2023 have been tested. In this research, from four deep learning algorithms (including long short term memory (LSTM), support vector machine (SVM), convolutional neural network (CNN) and recurrent neural network (RNN)) and also to select the final research variables from the method The average comparison test of two samples is used to create the model. The results of deep learning algorithms show that the overall accuracy of LSTM, SVM, CNN and RNN algorithms is 98.6%, 88.4%, 81.3% and 94.4% respectively, which shows that the LSTM algorithm has the best performance and the CNN algorithm has the worst performance in detecting companies with fraudulent financial statements. In other words, the results show the efficiency of long short term memory (LSTM) algorithm compared to other deep learning algorithms. Therefore, in listed companies on the Tehran Stock Exchange, the LSTM algorithm provides the most efficient model for detection of fraudulent financial statements. The findings of the present research can provide useful information for shareholders, creditors, analysts and other participants in the capital market in improving the prediction of fraudulent financial statements and reducing errors in making decisions based on financial statement information, better evaluating the corporate performance based on provide information, adopt optimal trading strategies and identify suitable opportunities for buying and selling stocks based on financial statement information.

Keywords: Fraudulent Financial Statements, Fraud Detection, Deep Learning, Long Short-Term Memory Algorithm
JEL classification: M41, M42, G32

[1] Department of Economics, Central Tehran Branch, Islamic Azad University, Tehran, Iran
javad.sharafkhani@iau.ac.ir
[2] Department of Economics, Central Tehran Branch, Islamic Azad University, Tehran, Iran. (Corresponding author) .bidari@ut.ac.ir