ارائه رویکردی مبتنی بر یادگیری عمیق برای کشف کلاه‌برداری در سرویس‌های پرداخت مالی

دوفصلنامه علمی

مدیریت اطلاعات

دوره ۵، شماره ۱

بهار و تابستان ۱۳۹۸

امیر حسین صدیقی

استادیار گروه سیستم‌های اطلاعاتی، پژوهشکده فناوری اطلاعات، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، تهران، ایران[1]

آرمان ساجدی‌نژاد

استادیار گروه مدیریت فناوری اطلاعات، پژوهشکده فناوری اطلاعات، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، تهران، ایران

چکیده: کشف خودکار کلاه‌برداری در سرویس‌های پرداخت مالی یکی از موضوعاتی است که با توجه به استفاده روزافزون از این نوع سرویس‌ها و افزایش حجم نقل و انتقالات مالی انجام‌شده در سیستم‌های بانکی از اهمیت بالایی برخوردار گشته است. بدین منظور نیازمند سیستمی هوشمند هستیم که بتواند با استفاده از ویژگی‌های مختلف یک تراکنش مالی، قانونی یا غیرقانونی بودن آن را به‌صورت بلادرنگ و با دقت قابل قبولی تشخیص دهد. برای طراحی چنین سیستمی در این مقاله از الگوریتمی مبتنی بر یادگیری عمیق بهره گرفته می‌شود. پس از تشریح الگوریتم پیشنهادی، کارایی آن با استفاده از یک مجموعه از تراکنش‌های مالی واقعی ارزیابی می‌گردد که به‌عنوان مجموعه داده معیار در پیشینه پژوهش شناخته می‌شود. سپس با استفاده از ملاک‌هایی نظیر صحت، دقت، معیار F، حساسیت و منحنی دقت- یادآوری مقایسه‌ای بین الگوریتم پیشنهادی با دو الگوریتم نزدیک‌ترین همسایگی و ماشین بردار پشتیبان صورت می‌گیرد. نتایج محاسباتی ضمن تائید کارایی الگوریتم پیشنهادی در مجموعه داده معیار، حاکی از صحت ۹۶ درصدی و دقت ۹۸ درصدی آن است.

کلیدواژه‌ها: فناوری اطلاعات، کشف کلاه‌برداری، یادگیری عمیق، یادگیری ماشینی، هوش مصنوعی.

۱. نویسنده مسئول: Seddighi@irandoc.ac.ir

۱. مقدمه

استفاده از سرویس‌های پرداخت مالی به بخشی جدایی‌ناپذیر از زندگی مدرن تبدیل‌شده است. این سرویس‌ها نحوه تعامل ما با دنیای پیرامونمان را متحول کرده و فرصت‌های جدیدی برای توسعه کسب‌وکارها به وجود آورده‌اند. بااین‌حال نباید از تهدیدات آن‌ها غافل بود. جرائم سایبری یکی از نگرانی‌های جدی در این حوزه است که در قالب کلاه‌برداری و سوءاستفاده از سرویس‌های پرداخت مالی، زیان‌های بسیاری را به طیف گسترده‌ای از ذینفعان شامل خریداران، فروشندگان، سازمان‌های مالی و بانک‌ها وارد می‌نماید.

با توجه به اهمیت موضوع، روش‌های مختلفی در ادبیات برای کشف کلاه‌برداری پیشنهاد شده‌اند. در رویکردهای سنتی برای کشف کلاه‌برداری نوعاً از روش‌های مبتنی بر قوانین[1] استفاده می‌شود. این بدان معناست که برای تشخیص کلاه‌برداری در یک تراکنش باید قوانینی به‌صورت دستی و از پیش، تعیین گردند. چنین سیستم‌هایی انعطاف‌پذیر نبوده و کلاه‌بردارها به‌سرعت راه‌هایی برای دور زدن آن‌ها پیدا می‌کنند. جایگزین مدرن این روش‌ها، استفاده از رویکردهای مبتنی بر یادگیری ماشینی است (Kou et al. 2004). در این روش‌ها ابتدا داده‌های عظیمی از تراکنش‌های بر خط جمع‌آوری و برچسب‌زنی می‌شوند. سپس ویژگی‌های اصلی این دادگان برای متمایزسازی تراکنش‌های قانونی از غیرقانونی استخراج می‌شوند. در گام بعد و با کمک الگوریتم‌های کلاس‌بندی با نظارت، این ویژگی‌ها آموخته‌شده و برای پیش‌بینی کلاه‌برداری در تراکنش‌های آینده به کار گرفته می‌شوند. بدین ترتیب این روش‌ها با استفاده از یک کلاس‌بندی باینری تراکنش‌های قانونی را از تراکنش‌های غیرقانونی و یا مشکوک متمایز می‌سازند.

ازآنجاکه نوعاً تعداد تراکنش‌های غیرقانونی ثبت‌شده بسیار کمتر از تعداد تراکنش‌های قانونی هستند، عملاً با یک عدم توازن شدید در نسبت داده‌های غیرقانونی به کل داده‌ها مواجه خواهیم بود که این امر کاربرد روش‌های کلاس‌بندی با نظارت را دچار مشکل می‌نماید. به‌عبارت‌دیگر هنگامی‌که داده‌های آموزش نامتوازن هستند، یک کلاس (تراکنش‌های قانونی) به‌طور نسبی نمود بیشتری از کلاس دیگر خواهد داشت به‌طوری‌که اگر این نسبت بسیار بالا باشد می‌تواند موجب آن شود که الگوریتم یادگیری، داده‌های کلاس کم‌جمعیت را به‌عنوان نویز در نظر گرفته و تمام داده‌ها را به‌عنوان نمونه‌هایی از تراکنش‌های قانونی کلاس‌بندی نماید (Japkowicz and Stephen 2002). ازاین‌رو دیده می‌شود که الگوریتم‌های یادگیری ماشینی که نوعاً به دنبال بیشینه کردن صحت[2] هستند در داده‌های نامتوازن جواب مناسبی به دست نمی‌دهند (He and Garcia 2009).

با توجه به آنچه گفته شد در این مقاله قصد داریم تا با استفاده از روش یادگیری عمیق با نظارت به کشف کلاه‌برداری در سرویس‌های پرداخت مالی بپردازیم. یادگیری عمیق یکی از روش‌های یادگیری ماشینی است که در سال‌های اخیر عملکرد خوبی را در حوزه‌های مختلفی همچون تشخیص گفتار و تشخیص تصاویر از خود نشان داده است. یادگیری عمیق مدل‌های محاسباتی را قادر می‌سازد تا با استفاده از چندین لایه پردازشی، اقدام به یادگیری دادگان نمایند (LeCun, Bengio, and Hinton 2015).

1.Rule-based
2.Accuracy

ادامه مقاله به این شرح سازمان‌یافته است. ابتدا در بخش دوم پیشینه پژوهش را بررسی خواهیم کرد. سپس در بخش سوم به معرفی الگوریتم یادگیری عمیق پیشنهادی خواهیم پرداخت. بخش چهارم عملکرد الگوریتم پیشنهادی را در یک مجموعه داده معیار مورد بررسی قرار داده و نتایج عددی حاصل را با دو الگوریتم برتر در این حوزه مقایسه می‌کند. بخش پنجم به بحث در نتایج اختصاص یافته است. درنهایت مقاله در بخش آخر نتیجه‌گیری می‌شود.

۲. پیشینه پژوهش

کشف کلاه‌برداری در تراکنش‌ها و سرویس‌های پرداخت مالی توجه زیادی را از سوی جامعه دانشگاهی به خود جلب کرده است. مقاله «بکر» یکی از اولین مطالعات در این حوزه با استفاده از روش‌های مصورسازی داده و یادگیری با نظارت است (Becker 1997). «جنسن» مشکلات فنی مربوط به عدم توازن کلاس‌ها در کشف کلاه‌برداری را مورد بحث و بررسی قرار داده است (Jensen 1997).

«شرمن» با کمک بررسی تراکنش‌های برچسب‌دار، یک راهکار نظارتی پیش‌گویانه پیشنهاد می‌دهد که به تعیین مشخصات تراکنش‌های غیرقانونی نوعی می‌پردازد (Sherman 2002). برخی پژوهشگران از شبکه‌های عصبی مصنوعی برای تعیین موارد کلاه‌برداری در حجم وسیعی از دادگان بهره گرفته‌اند. «قوش» و «ریلی» با استفاده از یک شبکه عصبی سه لایه اقدام به شناسایی کلاه‌برداری در یک مطالعه موردی کردند. آن‌ها موفق شدند که با عملکرد بهتر الگوریتم پیشنهادی خود نسبت به سیستم مبتنی بر قوانین موجود در بانک موردمطالعه، کاهشی بین ۲۰ تا ۴۰ درصد در کل زیان ناشی از تراکنش‌های غیرقانونی حاصل کنند (Ghosh and Reilly 1994). «آلسکروف»، «فیزلبن» و «رائو» سیستمی برای کاوش پایگاه داده با استفاده از یک شبکه عصبی سه لایه پیشنهاد دادند. آن‌ها موفق شدند تا با این الگوریتم تراکنش‌های غیرقانونی را با نرخ ۸۵ درصد کشف کنند (Aleskerov, Freisleben, and Rao 1997). در مطالعه دیگری «سیده»، «ژانگ» و «پن» یک شبکه عصبی موازی و پنج لایه را برای کشف تراکنش‌های غیرقانونی توسعه دادند و موفق شدند با نرخ ۷۵ درصد این نوع تراکنش‌ها را شناسایی کنند (Syeda, Zhang, and Pan 2002).

برخی دیگر از نویسندگان از روش درخت تصمیم برای ارائه راهکارهایی مبتنی بر قوانین در این حوزه استفاده کرده‌اند. «روست» و دیگران سیستمی دومرحله‌ای و مبتنی بر الگوریتم C4.5 برای تولید قوانین پیشنهاد دادند که توانست در کشف کلاه‌برداری به صحت در حدود ۹۰ درصد دست یابد (Rosset et al. 1999). «شائو»، «ژائو» و «چانگ» یک مدل چهار مرحله‌ای برای تشخیص کلاه‌برداری توسعه دادند (Shao, Zhao, and Chang 2002). آن‌ها در این مدل ابتدا قوانین تشخیص کلاه‌برداری را با استفاده از الگوریتم درخت تصمیم استخراج کرده و در مرحله بعد خبرگان این قوانین را بررسی، تائید و یا اصلاح می‌کنند. درنهایت این مدل بر اساس این قوانین ریسکی را به هر مورد تحت بررسی اختصاص می‌دهد. این روش در مثال‌های عددی تنها توانست کمی بیش از ۱۴ درصد صحت داشته باشد. نویسندگان دلیل این نتیجه را عدم توازن بیش‌ازاندازه نمونه‌های غیرقانونی در مقابل نمونه‌های قانونی بیان کرده‌اند.

«ماس» و دیگران مقایسه‌ای را بین راهبردهای با نظارت در کشف کلاه‌برداری بر مبنای شبکه‌های عصبی و شبکه‌های بیز ارائه داده‌اند (Maes et al. 2002). آن‌ها با استفاده از یک مجموعه داده واقعی به مقایسه این دو شبکه می‌پردازند. مطابق نتایج عددی به‌دست‌آمده نرخ تشخیص کلاه‌برداری برای شبکه عصبی و شبکه بیز به ترتیب برابر با ۷۰ و ۷۴ درصد گزارش شده است. بعلاوه زمان آموزش شبکه بیز سریع‌تر بوده و بدین ترتیب نویسندگان عملکرد شبکه بیز را بهتر از شبکه عصبی سه لایه مورداستفاده در این مطالعه ارزیابی کرده‌اند.

«کیم» و دیگران راه‌حلی مبتنی بر ماشین بردار پشتیبان را برای کشف کلاه‌برداری در شبکه‌های مخابراتی پیشنهاد می‌دهند (Kim et al. 2003). آن‌ها از ترکیب خروجی چندین مدل آموزش داده‌شده در حالت‌های مختلف استفاده کردند و موفق شدند در کشف کلاه‌برداری به نرخی در حدود ۹۷ درصد دست یابند. طبق یافته‌های آن‌ها ماشین بردار پشتیبان با کرنل تابع پایه شعاعی (RBF)[1] عملکرد بهتری در شناسایی موارد غیرقانونی دارد.

استفاده از روش بیز ساده در پژوهش «ویانه»، «دریگ» و «دِدن» و برای کشف کلاه‌برداری در درخواست‌های بیمه خودرو موردتوجه قرارگرفته است (Viaene, Derrig, and Dedene 2004). آن‌ها موفق شدند با تقویت روش بیز ساده به نرخ ۸۴ درصدی در شناسایی موارد غیرقانونی برسند.

«آویمی»، «آدتونمبی» و «الووادر» در یک تحلیل مقایسه‌ای به بررسی عملکرد روش‌های یادگیری ماشینی مختلف برای تشخیص کلاه‌برداری در تراکنش‌های کارت‌های اعتباری پرداختند (Awoyemi, Adetunmbi, and Oluwadare 2017). آن‌ها عملکرد الگوریتم نزدیک‌ترین همسایگی را بهتر از سایر الگوریتم‌ها بر روی مجموعه داده CCFD[2] ارزیابی کردند. این دادگان شامل مجموعه‌ای از تراکنش‌های برچسب‌زنی شده و واقعی مربوط به سیستم پرداخت مالی دسته‌ای از مشتریان اروپایی است که به‌صورت آزاد در دسترس عموم قرارگرفته و نوعاً به‌عنوان مجموعه داده معیار برای بررسی کارایی الگوریتم‌ها در این حوزه مورداستفاده قرار می‌گیرد (Dal Pozzolo et al. 2015).

۳. روش پژوهش

در این پژوهش از روش مطالعات کتابخانه‌ای برای طراحی الگوریتمی استفاده‌شده است که قادر به شناسایی خودکار تراکنش‌های غیرقانونی در یک سرویس پرداخت مالی است. در این راستا کشف کلاه‌برداری در تراکنش‌ها به‌صورت یک مسئله کلاس‌بندی باینری فرمول‌بندی می‌شود که در آن برداری از ویژگی‌ها و برچسب کلاس به هر یک از تراکنش‌ها متناظر می‌گردد. نوعاً مجموعه دادگان مورداستفاده در چنین مطالعاتی به‌شدت نامتوازن هستند، زیرا تراکنش‌های غیرقانونی تنها بخش کوچکی از کل تراکنش‌ها را به خود اختصاص می‌دهند. بااین‌حال تمرکز اصلی بر کلاس با جمعیت کمتر (تراکنش‌های غیرقانونی) معطوف

1.Radial Basis Function
2.Credit Card Fraud Detection Dataset

است. این امر نیاز به مدیریت صحیح مجموعه دادگان را بیش‌ازپیش مشخص می‌سازد. در این مقاله برای فائق آمدن بر این مشکل از روش نمونه‌برداری مجدد[1] استفاده خواهد شد.

با توجه به اینکه کشف تراکنش‌های غیرقانونی هدف اصلی در این پژوهش است، لذا از نمونه‌برداری محدود[2] برای ایجاد توازن بین دو کلاس مختلف استفاده خواهد شد (Drummond and Holte 2003). در این روش از کلاس با جمعیت بیش‌تر به‌صورت محدود نمونه‌برداری می‌گردد تا نسبت نهایی هر دو کلاس با یکدیگر برابر گردد. درنهایت مجموعه داده جدید به‌عنوان مجموعه داده معیار برای آموزش و آزمون الگوریتم‌های ماشینی مختلف مورداستفاده قرار خواهد گرفت. در ادامه ضمن معرفی روش یادگیری عمیق به تشریح الگوریتم پیشنهادی می‌پردازیم.

الگوریتم پیشنهادی

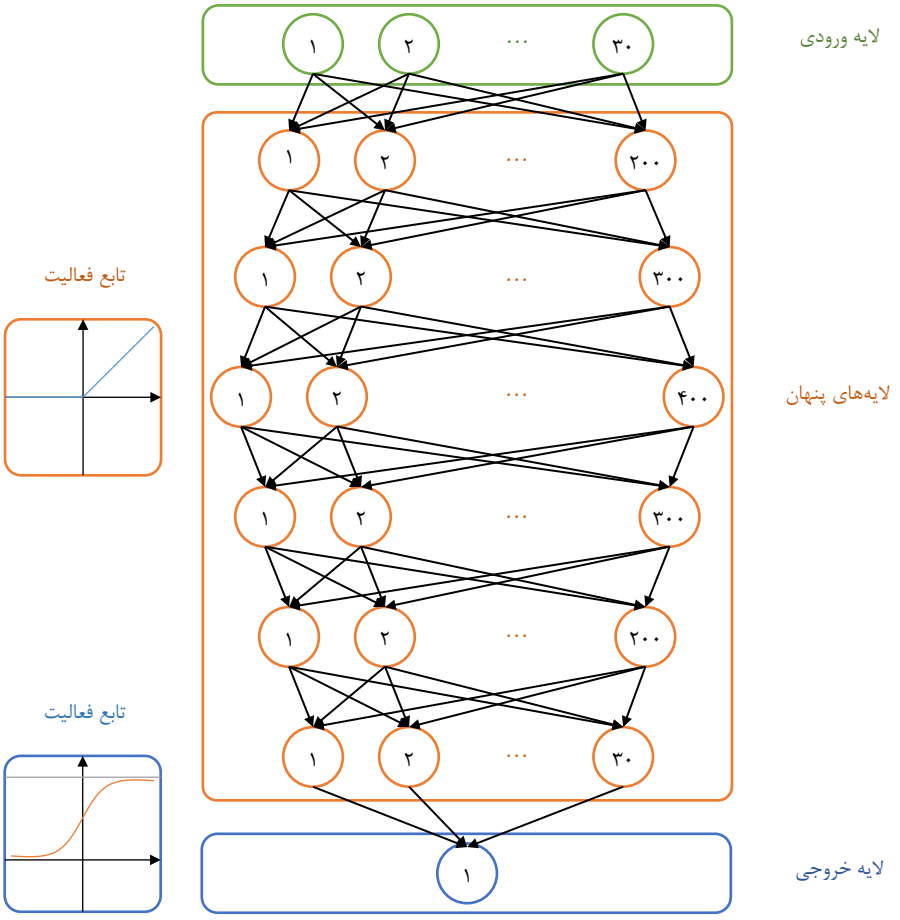
یادگیری عمیق به‌عنوان یکی از روش‌های یادگیری ماشینی به دنبال حل مسائلی است که نوعاً دیگر روش‌های یادگیری همچون ماشین بردار پشتیبان به دلیل ماهیت معماری کم‌عمق خود برای حل آن‌ها با مشکل مواجه هستند. در این روش با بهره‌گیری از یک معماری چندلایه از پردازشگرهای غیرخطی، مجموعه‌ای از ویژگی‌ها از دادگان آموزش استخراج می‌شوند که ازنظر آماری استوار هستند (Arnold et al. 2011).

به‌طورکلی یکی از مشکلات اصلی در یادگیری ماشینی انتخاب یک فضای ویژگی مناسب است به‌نحوی‌که داده‌های ورودی خصوصیات مطلوب برای حل یک مسئله خاص را داشته باشند. برای مثال در کلاس‌بندی باینری با نظارت، اغلب نیاز است تا دو کلاس با کمک یک ابرصفحه از یکدیگر جدا شوند. در حالتی که چنین خصوصیتی به‌صورت مستقیم در فضای ورودی قابل حصول نباشد، فرض می‌شود که بتوان این دادگان را به یک فضای ویژگی میانی نگاشت کرد که در آن کلاس‌ها به‌صورت خطی جدایی‌پذیر باشند. این فضای میانی می‌تواند یا به‌صورت مستقیم و با ویژگی‌های انتخاب‌شده به‌صورت دستی مشخص گردد، یا به‌صورت غیرمستقیم و با یک تابع کرنل تعریف شود و یا به‌صورت خودکار یادگیری شود. طراحی فضای ویژگی در هر دو حالت اول بر عهده کاربر است. این امر می‌تواند ازنظر زمان محاسباتی و دانش موردنیاز هزینه‌های بالایی را در پی داشته باشد به‌ویژه هنگامی‌که فضای ورودی از ابعاد بالایی برخوردار است. در حالت سوم، یادگیری خودکار ویژگی‌ها با معماری‌های عمیق (معماری‌های متشکل از چندین لایه از پردازشگرهای غیرخطی) می‌تواند به‌عنوان یک گزینه قابل‌قبول در نظر گرفته شود. به‌عبارت‌دیگر در یادگیری عمیق، لایه‌های مختلف از پردازشگرها سعی می‌کنند تا با نگاشت داده، آن را در سطحی بالاتر و کمی انتزاعی‌تر بازنمایی کنند. با ترکیب تعداد کافی از چنین نگاشت‌هایی، توابع بسیار پیچیده نیز قابلیت یادگیری خواهند داشت. به‌طور خاص در مسئله کلاس‌بندی، لایه‌های بازنمایی بالاتر، جنبه‌هایی از داده ورودی را برجسته می‌کنند که برای متمایز ساختن کلاس‌ها و حذف موارد پرت مهم هستند (LeCun, Bengio, and Hinton 2015).

1.Resampling
2.Under-Sampling

بدین ترتیب الگوریتم پیشنهادی با استفاده از روش یادگیری عمیق و به‌صورت یک شبکه عصبی مصنوعی عمیق در نظر گرفته می‌شود که از شش لایه پنهان تشکیل شده است (Hagan et al. 2014). شکل یک معماری عمیق این الگوریتم را نشان می‌دهد. برای تابع فعالیت نورون‌ها در لایه‌های پنهان از تابع یک‌سوساز خطی[1] (ReLU) استفاده‌شده است و نورون‌های لایه خروجی نیز از تابع سیگموئید به‌عنوان تابع فعالیت خود بهره می‌برند و آستانه تصمیم در لایه خروجی برابر با ۰/۵ در نظر گرفته‌شده است (Nair and Hinton 2010). به‌علاوه وزن اولیه نورون‌ها از توزیع تصادفی یکنواخت پیروی می‌کند. به‌منظور جلوگیری از بیش برازش شبکه عمیق پیشنهادی از راهبرد حذف برای آموزش این شبکه استفاده‌شده است (Srivastava et al. 2014). در این راهبرد، در هر بار به‌روزرسانی وزن‌ها، بخشی از نورون‌های هر لایه پنهان به‌صورت تصادفی انتخاب‌شده و حذف می‌گردند. باید توجه کرد که این حذف موقتی بوده و بعد از پایان مرحله آموزش، تمامی نورون‌ها در شبکه حاضر خواهند بود و تنها وزنشان به صورتی اصلاح می‌گردد که شبکه نهایی به‌صورت تقریبی خروجی معادل با شبکه آموزش‌یافته را به دست دهد. برای توضیحات بیشتر درباره این روش به مقاله «سیرواستاوا» و دیگران مراجعه کنید (Srivastava et al. 2014). درنهایت برای آموزش شبکه عمیق پیشنهادی از روش گرادیان کاهشی تصادفی[2] (SGD) بهره گرفته می‌شود (Bottou 2012).

1.Rectified Linear Unit
2.Stochastic Gradient Descent



شکل ۱- معماری عمیق الگوریتم پیشنهادی

٤. تجزیه‌وتحلیل داده‌ها

در این بخش عملکرد شبکه عمیق پیشنهادی را بررسی و ارزیابی خواهیم کرد. بدین منظور از مجموعه داده معیار CCFD بهره خواهیم گرفت که از آدرس «https://datahub.io/machine-learning/creditcard» قابل‌دسترس است (Dal Pozzolo et al. 2015). این مجموعه داده شامل ۲۸۴۸۰۷ تراکنش می‌شود که در طی دو روز در سپتامبر ۲۰۱۳ توسط مشتریان اروپایی به ثبت رسیده‌اند. دراین‌بین، تعداد ۴۹۲ تراکنش وجود دارند که برچسب غیرقانونی داشته و درمجموع ۰/۱۷۲ درصد از کل تراکنش‌ها در مجموعه داده معیار را تشکیل می‌دهند که حاکی از عدم توازن شدید در این داده‌هاست. برای حفظ محرمانگی اطلاعات، ویژگی‌های اصلی و اطلاعات پیش‌زمینه درباره آن‌ها در اختیار قرار داده نشده‌اند. فقط ویژگی‌های عددی این مجموعه داده و آن‌هم بعد از ناشناس سازی در دسترس هستند. این ویژگی‌های

عددی که با اعمال تحلیل عناصر اصلی (PCA)[1] بر روی ویژگی‌های اصلی مجموعه داده معیار به‌دست‌آمده‌اند با برچسب‌های V1 تا V28 نشانه‌گذاری شده‌اند. تنها ویژگی‌هایی که با PCA تغییرنیافته‌اند، زمان[2] و میزان تراکنش[3] هستند. ویژگی زمان نشانگر تعداد ثانیه‌های سپری‌شده بین هر تراکنش با تراکنش اول در مجموعه داده معیار است. ویژگی کلاس[4] بیان‌کننده متغیر پاسخ یا برچسب هر تراکنش است و درصورتی‌که تراکنش غیرقانونی باشد مقدار یک را به خود می‌گیرد و در صورت قانونی بودن مقدار صفر را خواهد داشت.

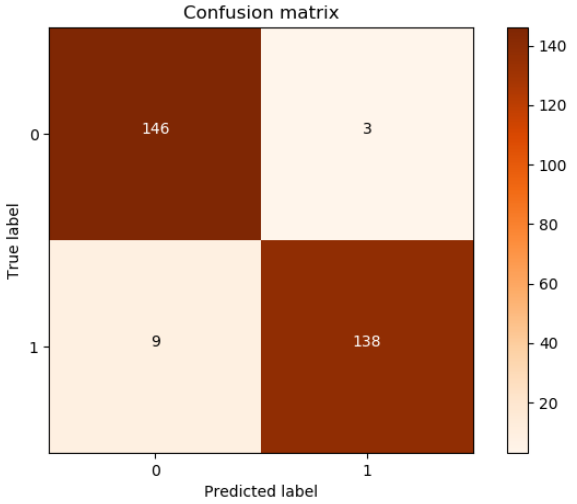
لازم به ذکر است که برای انجام آزمایش‌های عددی ارائه‌شده در این بخش از کتابخانه‌های scikit-learn و tensorflow در زبان پایتون کمک گرفته‌شده است (Pedregosa et al. 2011; Abadi et al. 2016).

زمان آموزش شبکه عمیق پیشنهادی بر روی کامپیوتری با یک پردازنده چهار هسته‌ای باقدرت پردازش ۳/۳ گیگاهرتز و چهار گیگابایت رم در حدود یک دقیقه است. بعلاوه خوانندگان می‌توانند برای دسترسی به شبکه آموزش داده‌شده نهایی با نویسندگان مقاله مکاتبه نمایند.

برای آماده‌سازی دادگان، ابتدا تمامی ۳۰ ویژگی یادشده شامل ویژگی‌های V1 تا V28، زمان و میزان تراکنش به‌صورت مجزا نرمال می‌شوند. بدین ترتیب که میانگین هر ویژگی از هر عنصر آن کاسته شده و حاصل بر انحراف استاندارد آن ویژگی تقسیم می‌شود. سپس با استفاده از نمونه‌برداری محدود یک مجموعه داده متوازن شده، متشکل از ۹۸۴ تراکنش به دست می‌آید. بدین منظور ابتدا ۴۹۲ تراکنش قانونی به‌تصادف از مجموعه تراکنش‌های قانونی انتخاب می‌شوند و سپس با ۴۹۲ تراکنش غیرقانونی موجود در مجموعه داده معیار ترکیب می‌گردند تا مجموعه داده‌ای متوازن با نسبت ۵۰ درصد از دو نوع تراکنش حاصل گردد. در ادامه این مجموعه داده متوازن شده به‌صورت تصادفی و با نسبت ۷ به ۳ به مجموعه دادگان یادگیری و آزمون تقسیم‌بندی می‌گردد. بدین ترتیب تعداد ۶۸۸ تراکنش برای یادگیری و تعداد ۲۹۶ تراکنش برای آزمون الگوریتم پیشنهادی مورداستفاده قرار خواهند گرفت.

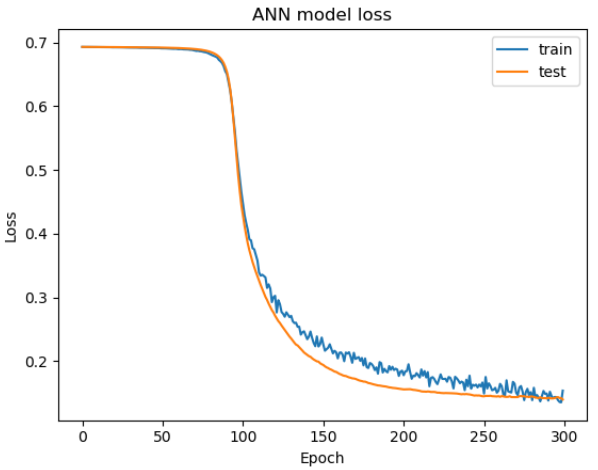
شکل دو عملکرد الگوریتم پیشنهادی را برحسب تعداد مثبت‌های درست، منفی‌های درست، مثبت‌های نادرست و منفی‌های نادرست نشان می‌دهد. در این شکل محور افقی نشانگر برچسب‌های پیش‌بینی‌شده توسط الگوریتم پیشنهادی و محور عمودی معرف برچسب اصلی دادگان آزمون است. مطابق نتایج به‌دست‌آمده میزان صحت این الگوریتم ۰/۹۵۹ و میزان دقت آن ۰/۹۷۹ محاسبه می‌گردد که نشان از عملکرد قابل‌قبول این الگوریتم دارد.

1.Principal Components Analysis
2.Time
3.Amount
4.Class



شکل ۲- ماتریس درهم‌ریختگی الگوریتم پیشنهادی

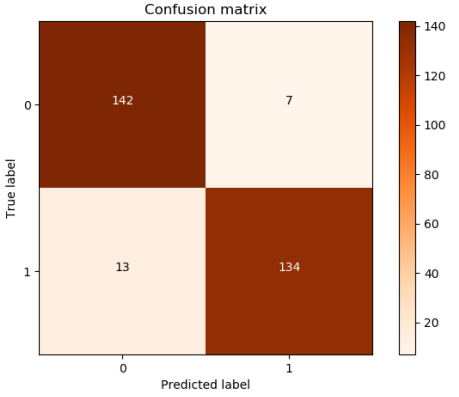
شکل سه نمودار تابع زیان[1] را برای الگوریتم پیشنهادی نشان می‌دهد. تابع زیان بیانگر خطای یادگیری است و در طول فرآیند آموزش باید کاهش یابد. آموزش شبکه هنگامی متوقف می‌شود که میزان این تابع از حد مشخصی کمتر شود. مطابق این نمودار می‌توان نتیجه گرفت که آموزش شبکه عمیق پیشنهادی بر روی دادگان یادگیری تا حد مناسبی انجام‌شده است. از طرف دیگر خطای کلاس‌بندی بر روی دادگان آزمون نیز کاهش مشابهی را در طول فرآیند یادگیری از خود نشان می‌دهد.



1.Loss Function

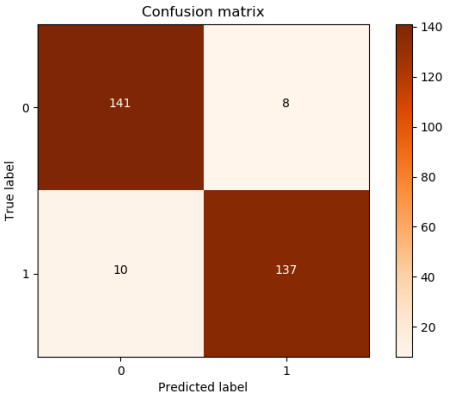
شکل ۳- نمودار تابع زیان در حین فرآیند یادگیری الگوریتم پیشنهادی

برای ارزیابی بهتر عملکرد الگوریتم پیشنهادی، نتایج حاصل از آن را با الگوریتم‌های نزدیک‌ترین همسایگی (KNN)[1] و ماشین بردار پشتیبان (SVM)[2] مقایسه خواهیم کرد. الگوریتم نزدیک‌ترین همسایگی در پیشینه پژوهش به‌عنوان الگوریتم برتر در مقایسه با سایر روش‌های یادگیری ماشینی برای کشف کلاه‌برداری در مجموعه داده معیار معرفی‌شده است (Awoyemi, Adetunmbi, and Oluwadare 2017). ماشین بردار پشتیبان نیز یک مدل یادگیری با نظارت است. این مدل به دنبال یافتن ابر صفحه‌ای در ابعاد بالا است که بتواند دادگان آموزش را به نحوی جدا نماید که دارای بیشترین فاصله از نزدیک‌ترین نقاط آموزش از هر کلاس باشد (Cortes and Vapnik 1995). ازآنجاکه فضای دادگان در این پژوهش غیرخطی است از کرنل تابع پایه شعاعی (RBF) برای ماشین بردار پشتیبان استفاده خواهیم کرد (Hofmann 2006). شکل‌های چهار و پنج به ترتیب عملکرد الگوریتم‌های نزدیک‌ترین همسایگی و ماشین بردار پشتیبان را با استفاده از ماتریس درهم‌ریختگی نشان می‌دهند.



شکل ۴- ماتریس درهم‌ریختگی الگوریتم نزدیک‌ترین همسایگی

1.K-Nearest Neighbors
2.Support Vector Machine



شکل ۵- ماتریس درهم‌ریختگی الگوریتم ماشین بردار پشتیبان

با مقایسه شکل‌های دو، چهار و پنج متوجه می‌شویم که الگوریتم پیشنهادی تنها ۹ تراکنش غیرقانونی را به‌صورت اشتباه تشخیص داده است درحالی‌که عملکرد الگوریتم‌های ماشین بردار پشتیبان و نزدیک‌ترین همسایگی پایین‌تر بوده و به ترتیب ۱۰ و ۱۳ تراکنش غیرقانونی را قانونی تشخیص داده‌اند. به‌علاوه الگوریتم پیشنهادی سه تراکنش قانونی را غیرقانونی تشخیص داده است که در مقایسه با هفت و هشت تراکنشی که الگوریتم‌های نزدیک‌ترین همسایگی و ماشین بردار پشتیبان اشتباه تشخیص داده‌اند، عملکرد بهتری است.

جدول یک عملکرد این سه الگوریتم را بر روی دادگان آزمون با استفاده از شاخص‌های حساسیت[1] یا یادآوری[2]، دقت[3]، معیار F و صحت نشان می‌دهد. حساسیت بیانگر توانایی کشف یک مورد تراکنش غیرقانونی است به‌شرط آنکه واقعاً غیرقانونی باشد. به‌عبارت‌دیگر حساسیت نسبت تراکنش‌های غیرقانونی درست تشخیص داده‌شده (تعداد مثبت درست) به کل تراکنش‌های غیرقانونی (مجموع مثبت درست و منفی نادرست) است. دقت نسبت تراکنش‌های غیرقانونی درست تشخیص داده‌شده (تعداد مثبت درست) به کل تراکنش‌هایی است که غیرقانونی تشخیص داده‌شده‌اند (مجموع مثبت درست و مثبت نادرست). معیار F، میانگین هارمونیک حساسیت و دقت است و صحت نسبت پیش‌بینی‌هایی است که درست تشخیص داده‌شده‌اند. همان‌طور که در این جدول دیده می‌شود، عملکرد الگوریتم پیشنهادی در تمامی معیارها مذکور بهتر از الگوریتم‌های نزدیک‌ترین همسایگی و ماشین بردار پشتیبان است.

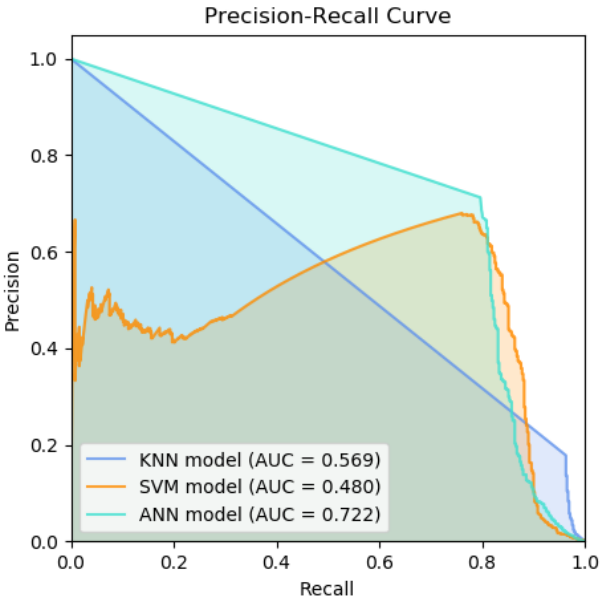
جدول ۱- مقایسه عملکرد الگوریتم‌های مختلف

الگوریتم‌ها	صحت	دقت	معیار F	حساسیت
الگوریتم پیشنهادی	۰/۹۵۹	۰/۹۷۹	۰/۹۵۸	۰/۹۳۹
الگوریتم نزدیک‌ترین همسایگی	۰/۹۳۲	۰/۹۵۰	۰/۹۳۱	۰/۹۱۲

1.Sensitivity
2.Recall
3.Precision

الگوریتم ماشین بردار پشتیبان	۰/۹۳۹	۰/۹۴۵	۰/۹۳۸	۰/۹۳۲

درنهایت شکل شش سعی دارد تا با استفاده از منحنی دقت- یادآوری (PRC)[1] عملکرد الگوریتم پیشنهادی را بر روی کل مجموعه دادگان نامتوازن در دسترس ارزیابی نماید. این منحنی نسبت بین دقت و یادآوری را در آستانه‌های مختلف نشان می‌دهد و در مواردی کاربرد دارد که نسبت جمعیت کلاس‌ها بسیار نامتوازن باشد. مطابق این شکل الگوریتم پیشنهادی با داشتن سطح زیر نمودار بیشتر، دارای عملکرد بهتری در مقایسه با دو الگوریتم دیگر است.



شکل ۶- منحنی دقت- یادآوری برای الگوریتم‌های مختلف روی کل مجموعه دادگان

۵. بحث در نتایج

یکی از موارد مهم در ارزیابی الگوریتم‌های کشف کلاه‌برداری تعداد تراکنش‌های غیرقانونی است که قانونی تشخیص داده می‌شوند (منفی‌های نادرست). به‌عبارت‌دیگر در هنگام کشف کلاه‌برداری در سرویس‌های پرداخت مالی، آنچه از اهمیت بیشتری برخوردار است و می‌تواند برای موسسه‌های مالی، بانک‌ها و مشتریانشان هزینه سنگینی داشته باشد، تراکنش‌هایی است که غیرقانونی هستند اما توسط سیستم قانونی تشخیص داده می‌شوند؛ بنابراین نوعاً در این حوزه پژوهشی به دنبال روش‌هایی هستیم که تا جای ممکن منفی‌های نادرست کمتری را به دست دهند. بااینکه تعداد تراکنش‌های قانونی که غیرقانونی تشخیص داده می‌شوند (مثبت‌های نادرست) نیز می‌توانند تا حدی بر اعتبار سازمان‌های مالی تأثیرگذار باشند اما در

1.Precision-Recall Curve

مقایسه با زیان ناشی از کلاه‌برداری‌ها عمدتاً از آن صرف‌نظر می‌شود. به هر رو روش پیشنهادی هم ازنظر تعداد منفی‌های نادرست و هم ازنظر تعداد مثبت‌های نادرست در مقایسه با دو الگوریتم دیگر دارای برتری است.

با نگاهی به نتایج به‌دست‌آمده از الگوریتم پیشنهادی ممکن است این موضوع به ذهن خطور کند که شاید بتوان با بررسی نمونه‌هایی که به‌اشتباه تشخیص داده‌شده‌اند به دلایل و شرایط منجر به آن پی برد. بااینکه این امر می‌تواند در طراحی بهتر شبکه و یا انتخاب پیش‌پردازش‌های مناسب برای دادگان حائز اهمیت باشد اما به چند دلیل در این پژوهش قابل حصول نیست. ازجمله این دلایل می‌توان به ابعاد بالای ویژگی‌های ورودی (۳۰ ویژگی)، عدم آگاهی از معنای ضمنی این ویژگی‌ها به دلیل ناشناس سازی داده‌ها و تعداد نسبتاً زیاد لایه‌های شبکه پیشنهادی اشاره کرد. بدین ترتیب یافتن علل تشخیص اشتباه در این شبکه بدون اطلاعات پیش‌زمینه‌ای درباره دادگان و ویژگی‌های آن‌ها کار دشواری است. بااین‌حال به‌صورت کلی می‌توان گفت که با افزایش میزان نمونه‌های مربوط به تراکنش‌های غیرقانونی و آموزش دوباره شبکه با حجم بیشتری از دادگان، شبکه قادر خواهد بود تا الگوهای کلاه‌برداری را که تاکنون تشخیص نداده بازشناسی کند و بدین ترتیب عملکرد خود را در شناسایی موارد موجود بهبود بخشد.

از طرف دیگر با به‌روز نگاه‌داشتن داده‌های ورودی، شبکه می‌تواند الگوهای جدید کلاه‌برداری را نیز یاد گرفته و تشخیص دهد. این امر تا حد قابل‌توجهی از مشکلات موجود در روش‌های سنتی مقابله با کلاه‌برداری نظیر روش‌های مبتنی بر قاعده که عمدتاً نیازمند افزودن و یا اصلاح دستی قواعد هستند خواهد کاست.

الگوریتم پیشنهادی ازنظر عملکرد بر روی کل مجموعه دادگان نتایج نسبتاً قابل قبولی ارائه می‌دهد (شکل ۶). باید خاطرنشان کرد که این نتیجه تنها با آموزش شبکه بر روی ۶۸۸ نمونه حاصل‌شده است و سپس مدل نهایی در یک مجموعه داده با بیش از ۲۸۴ هزار نمونه مورد آزمایش قرارگرفته است. این امر ناشی از توانایی بالای شبکه‌های عصبی عمیق در تعمیم‌پذیری الگوهای یادگیری شده در مقایسه با دو الگوریتم دیگر است. بااین‌حال می‌توان سطح عملکرد فعلی شبکه پیشنهادی را با افزایش تعداد تراکنش‌های غیرقانونی و استفاده از دادگان با حجم بزرگ‌تر در مرحله یادگیری ارتقاء داد.

حجم پایین نمونه مورداستفاده در این مقاله از محدودیت‌های پژوهش است که باید در پژوهش‌های آتی و با بهره‌گیری از مطالعه موردی در موسسه‌های مالی و بانک‌ها در جهت برطرف سازی آن و ارزیابی دقیق‌تر شبکه عمیق پیشنهادی گام برداشت. بااین‌حال می‌توان با توجه به کارایی قابل‌قبول الگوریتم پیشنهادی بر روی مجموعه دادگان معیار، استفاده از این روش را برای تشخیص خودکار کلاه‌برداری در سرویس‌های پرداخت مالی توصیه کرد.

۶. جمع‌بندی و نتیجه‌گیری

در این مقاله روشی مبتنی بر یادگیری عمیق برای کشف کلاه‌برداری در تراکنش‌ها و سرویس‌های پرداخت مالی پیشنهاد شد. سپس عملکرد روش پیشنهادی با استفاده از شاخص‌های گوناگون بر روی مجموعه‌ای از تراکنش‌های برچسب‌زنی شده و واقعی مربوط به سیستم پرداخت مالی دسته‌ای از مشتریان اروپایی مورد

بررسی و ارزیابی قرار گرفت. این مجموعه داده که به‌صورت آزاد در دسترس عموم است نوعاً به‌عنوان مجموعه داده معیار برای بررسی کارایی الگوریتم‌ها در این حوزه مورداستفاده قرار می‌گیرد.

نتایج این پژوهش نشان داد که الگوریتم پیشنهادی با ۹۶ درصد صحت، ۹۸ درصد دقت و ۹۴ درصد حساسیت در تشخیص کلاه‌برداری، در مقایسه با الگوریتم نزدیک‌ترین همسایگی به‌عنوان الگوریتم برتر مورداشاره در پیشینه موضوع و الگوریتم ماشین بردار پشتیبان کارایی بالاتری بر روی مجموعه داده معیار دارد.

بدین ترتیب می‌توان استفاده از الگوریتم‌های یادگیری عمیق را برای بهبود سیستم‌های تشخیص کلاه‌برداری در بانک‌ها و مؤسسات مالی پیشنهاد داد. نتایج مشابهی با بهره‌گیری از یادگیری عمیق برای تشخیص کلاه‌برداری‌های تلفنی از مشتریان توسط «ژنگ» و دیگران به‌دست‌آمده است (Zheng et al. 2018)؛ در این مطالعه دیده شد که دو بانک چینی با به‌کارگیری یادگیری عمیق موفق شدند تا در طی ۱۲ هفته در حدود ۱۰ میلیون یوآن خسارات ناشی از کلاه‌برداری را کاهش دهند.

یکی از چالش‌های اصلی در طراحی روش‌های یادگیری ماشینی در این حوزه، عدم توازن شدید دادگان مربوط به تراکنش‌های مالی است. این امر ناشی از میزان بسیار پایین تراکنش‌های غیرقانونی نسبت به تعداد کل تراکنش‌ها است. برای فائق آمدن بر این چالش از یک رویکرد مبتنی بر نمونه‌برداری مجدد کمک گرفته شد که توانست نتایج قابل قبولی را حاصل نماید. این یافته با نتایج حاصل از مطالعه «ژاپکویکز» و «استفن» همخوانی دارد که در آن استفاده از روش‌های نمونه‌برداری مجدد در شبکه‌های عصبی چندلایه قابل‌توجه ارزیابی‌شده بود (Japkowicz and Stephen 2002).

درنهایت به‌عنوان تحقیقات آتی می‌توان با پیاده‌سازی الگوریتم پیشنهادی در یک نمونه موردی در دنیای واقعی، ضمن ارتقای این روش، تصویر بهتری از چالش‌های پیش روی آن در بعد عملی به دست آورد.

فهرست منابع

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng. 2016. TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation (OSDI)*, 265-283.

Aleskerov, E., B. Freisleben, and B. Rao. 1997. CARDWATCH: a neural network based database mining system for credit card fraud detection. In *Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr)*, 220-226.

Arnold, L., S. Rebecchi, S. Chevallier, and H. Paugam-Moisy. 2011. An Introduction to Deep Learning. In *Proceedings of the European Symposium on Artificial Neural Networks (ESANN)*, 477-488.

Awoyemi, J. O., A. O. Adetunmbi, and S. A. Oluwadare. 2017. Credit card fraud detection using machine learning techniques: A comparative analysis. In *Proceedings of the International Conference on Computing Networking and Informatics (ICCNI)*, 1-9.

Becker, B. G. 1997. Using Mineset for Knowledge Discovery. *IEEE Computer Graphics and Applications* 17 (4): 75-78.

Bottou, L. 2012. Stochastic Gradient Descent Tricks. In *Neural Networks: Tricks of the Trade*. Lecture Notes in Computer Science. Ed. G. Montavon, G. B. Orr, and K.-R. Müller, 421-436. Berlin: Springer.

Cortes, C., and V. Vapnik. 1995. Support-Vector Networks. *Machine Learning* 20 (3): 273-297.

Dal Pozzolo, A., O. Caelen, R. A. Johnson, and G. Bontempi. 2015. Calibrating Probability with Undersampling for Unbalanced Classification. In *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, 159-166.

Drummond, C., and R. C. Holte. 2003. C4.5, Class Imbalance, and Cost Sensitivity: Why Under-Sampling Beats Over-Sampling. In *Proceedings of the Workshop on Learning from Imbalanced Datasets II*, 1-8.

Ghosh, S., and D. L. Reilly. 1994. Credit Card Fraud Detection with a Neural-Network. In *Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences*, 621-630.

Hagan, M. T., H. B. Demuth, M. H. Beale, and O. De Jesús. 2014. *Neural Network Design, 2nd edition*. USA: Martin Hagan.

He, H., and E. A. Garcia. 2009. Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering* 21 (9): 1263-1284.

Hofmann, M. 2006. Support Vector Machine - Kernel and The Kernel Trick. *An elaboration for the Hauptseminar*, Bamberg University, 1-16.

Japkowicz, N., and S. Stephen. 2002. The Class Imbalance Problem: a Systematic Study. *Intelligent Data Analysis* 6 (5): 429-449.

Jensen, D. 1997. Prospective Assessment of AI Technologies for Fraud Detection: A Case Study. In *Proceedings of the AAAI Workshop on AI Approaches to Fraud Detection and Risk Management*, 34-38.

Kim, H.-C., S. Pang, H.-M. Je, D. Kim, and S. Y. Bang. 2003. Constructing Support Vector Machine Ensemble. *Pattern Recognition* 36 (12): 2757-2767.

Kou, Y., C.-T. Lu, S. Sirwongwattana, and Y.-P. Huang. 2004. Survey of fraud detection techniques. In *Proceedings of the IEEE International Conference on Networking, Sensing and Control*, 749-754.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. *Nature* 521: 436-444.

Maes, S., K. Tuyls, B. Vanschoenwinkel, and B. Manderick. 2002. Credit Card Fraud Detection Using Bayesian and Neural Networks. In *Proceedings of the 1st International Naiso Congress on Neuro Fuzzy Technologies*, 261-270.

Nair, V., and G. E. Hinton. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML)*, 807-814.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12: 2825-2830.

Rosset, S., U. Murad, E. Neumann, Y. Idan, and G. Pinkas. 1999. Discovery of Fraud Rules for Telecommunications? Challenges and Solutions. In *Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 409-413.

Shao, H., H. Zhao, and G.-R. Chang. 2002. Applying Data Mining to Detect Fraud Behavior in Customs Declaration. In *Proceedings of the International Conference on Machine Learning and Cybernetics*, 1241-1244.

Sherman, E. 2002. Fighting Web Fraud. *Newsweek* 139 (23): 32B-32B.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15 (Jun): 1929-1958.

Syeda, M., Y.-Q. Zhang, and Y. Pan. 2002. Parallel Granular Neural Networks for Fast Credit Card Fraud Detection. In *Proceedings of the 2002 IEEE International Conference on Fuzzy Systems*, 572-577.

Viaene, S., R. A. Derrig, and G. Dedene. 2004. A Case Study of Applying Boosting Naive Bayes to Claim Fraud Diagnosis. *IEEE Transactions on Knowledge and Data Engineering* 16 (5): 612-620.

Zheng, Y.-J., X.-H. Zhou, W.-G. Sheng, Y. Xue, and S.-Y. Chen. 2018. Generative adversarial network based telecom fraud detection at the receiving bank. *Neural Networks* 102: 78-86.

A Deep Learning Approach to Fraud Detection in Financial Payment Services

Amir Hossein Seddighi
Assistant Prof., Information Technology Research Department, Iranian Research Institute for Information Science and Technology (IranDoc), Tehran, Iran[1]
Arman Sajedinejad
Assistant Prof., Information Technology Research Department, Iranian Research Institute for Information Science and Technology (IranDoc), Tehran, Iran

Abstract: Widespread use of financial payment services besides the increased volume of financial transfers carried out in banking systems, have resulted in an important growing trend of automatic fraud detection. In this regard, an intelligent system is needed that can determine the fraudulence or genuineness of a financial transaction in real-time, unquestionably with an acceptable precision using the different transaction features. In order to gain the benefits of the system, a deep learning algorithm is described and proposed in this paper. The performance of the proposed algorithm is evaluated using a set of real-world financial transactions, which is known as the standard dataset in the literature. Then, the proposed algorithm is compared with k-nearest neighbors and support vector machine algorithms using different metrics such as accuracy, precision, F-measure, sensitivity, and precision-recall curve. The computational results confirmed the efficiency of the proposed algorithm on the standard dataset with 96% accuracy and 98% precision.
Keywords: Artificial Intelligence, Deep Learning, Fraud Detection, Information Technology, Machine Learning.

1.Corresponding Author: Seddighi@irandoc.ac.ir