# بررسی کلان داده ها در تشخیص حملات در اینترنت با استفاده از هوش مصنوعی

**محمدرضا حسین زاده مقدم**

دانشجو دکتری، مهندسی کامپیوتر گرایش نرم افزار ، دانشگاه آزاد اسلامی واحد تهران جنوب ، تهران ، ایران

**چکیده**

امروزه هوش مصنوعی شبیه‌سازی فرآیندهای هوش انسانی توسط ماشین‌ها به ویژه سیستم‌های کامپیوتری است. به‌طور کلی، سیستم‌های هوش مصنوعی با دریافت مقادیر زیادی از داده‌های آموزشی برچسب‌گذاری‌شده، تجزیه و تحلیل داده‌ها برای همبستگی‌ها و الگوها و استفاده از این الگوها برای پیش‌بینی وضعیت‌های آینده کار می‌کنند و می‌تواند یاد بگیرد که تبادلات واقعی با افراد ایجاد کند، یا یک ابزار تشخیص تصویر می‌تواند با مرور میلیون‌ها مثال، شناسایی و توصیف اشیاء در تصاویر را بیاموزد. هوش مصنوعی این امکان را برای ماشین‌ها فراهم می‌کند تا با ورودی‌های جدید سازگار شوند و کارهایی شبیه به انسان را انجام دهند. بیشتر نمونه‌های هوش مصنوعی که امروزه درباره آن‌ها می‌شنوید به شدت به یادگیری عمیق و پردازش زبان طبیعی متکی هستند. با استفاده از این فناوری‌ها، به رایانه‌ها می‌توان برای انجام وظایف خاص با پردازش مقادیر زیادی داده و تشخیص الگوهای موجود در داده‌ها آموزش داد.

در این روش از یادگیری ماشین با نظارت استفاده می کنیم زیرا در یادگیری با نظارت ماشین با استفاده از داده‌های برچسب گذاری شده و داشتن جواب‌های درست را یاد میگیرد در این روش از الگوریتم SIEM برای فیلترکردن داده و استفاده در فضای سازمان را انجام می دهیم که می تواند در تشخیص حملات فیشینک به خوبی عمل کند.

**واژگان کلیدی:** الگوریتم SIEM، یادگیری ماشین با نظارت، هوش مصنوعی، داده های حجیم و اینترنت

**مقدمه**

در این مقاله در مورد تشخیص حملات در اینترنت در کلان داده ها [1] صحبت می کنیم، زیرا امروزه بدلیل حجم انبوهی از داده ها نمی توان با استفاده از ابزار های سنتی به ذخیره ، پردازش یا تجزیه و تحلیل اطلاعات پرداخت به همین دلیل باید با استفاده از ابزارهای هوشمند از جمله هوش تجاری که میلیون های منبع وجود دارد داده ها را با سرعت بالایی تولید می کند. این منابع داده ها را در سرتاسر جهان وجود دارد که نیاز به پردازش و تجزیه و تحلیل دارد تا بتوانیم از بروز حمله به سیستم های کامپیوتری جلوگیری کنیم این منابع که امروزه وجود دارد شبکه های اجتماعی همانند فیسبوک و شبکه های کامپیوتری می باشد که تمام افراد روزانه با این شبکه ها جهت ارسال تصاویر، فیلم ها و پیام ها و موارد دیگر در ارتباط هستند. داده ها همچنین در قالب های مختلفی از جمله داده های ساختار یافته ، نیمه ساختاریافته و غیر ساخت یافته بیان می شود می توان گفت چون در فضای اینترنت ما پردازش اطلاعات را انجام می دهیم دادهای بدون ساخت یافته داریم که باید همه داده ها را در یک صفحه معمولی اکسل داده ها را به عنوان داده های ساختار یافته طبقه بندی می شوند با یک قالب مشخص در مقابل ایمیل ها در بخش نیمه ساخت یافته قرار می گیرند و تصاویر و ویدیوهای افراد در زیر داده های بدون ساخت یافته قرار می گیرند. همه این داده ها که با یکدیگر ترکیب شده اند و کلان داده یا Big Data را تشکیل می دهند.

در این مقاله با تجزیه و تحلیل کلان داده ها می توانیم فرآیندی برای بدست آوردن درک عمیق مانند الگوی پنهان، همبستگی های ناشناخته، ترندهای بازار و ترجیحات کاربران، تجزیه و تحلیل کلان داده ها مزایای مختلفی از جمله برای تصمیم گیری بهتر و جلوگیری از فعالیت های متقلبانه و موارد دیگر در این میان استفاده کرد و همچنین داده تا زمانی که به اطلاعات و دانش مفیدی تبدیل نشود که بتواند به مدیریت در تصمیم گیری کمک کند بی معنی است برای این منظور ما چندین نرم افزار در کلان داده داریم که در این نرم افزار ها به ذخیره، تجزیه و تحلیل، گزارش گیری و انجام کارهای بیشتر با داده ها کمک می کند که در ادامه مقاله در بخش دوم ادبیات تحقیق، در بخش سوم روش اجرایی، بخش چهارم یافته ها (تجزیه و تحلیل) و در بخش آخر نتیجه گیری و پیشنهادات کارهای آتی را مورد بررسی قرار می دهیم.

**ادبیات تحقیق**

در این مقاله ( هانیه شامبیانی و همکاران،۱۴۰۱ ) در خصوص بهینه سازی عملکرد پردازش اطلاعات در زنجیره تامین مجازی مبتنی بر اینترنت اشیاء را مورد بررسی قرارداده اند و در این پژوهش امروزه، صنعت تولید با گسترش محدودیت‌های فیزیکی تجارت در سطح جهان، فناوری‌های مدرن اطلاعاتی را به‌منظور بهینه‌سازی روند تجارت و دستیابی به ادغام با شرکای زنجیرهٔ تأمین در پیش گرفته است که ازنظر جغرافیایی پراکنده‌اند. مدل‌های سنتی زنجیرهٔ تأمین، توجه اصلی را به بهینه‌سازی جریان‌های فیزیکی می‌دهد؛ با وجود این، اطمینان از اینکه واحدهای فیزیکی قابلیت پردازش اطلاعات مناسب را دارد نیز به همان اندازه مهم است. به این منظور در این پژوهش، بهینه‌سازی عملکرد پردازش اطلاعات در زنجیرهٔ تأمین مجازی حلقهٔ بسته، با هدف حداکثرسازی شود و سرعت پردازش اطلاعات، با در نظر گرفتن هزینه‌های مجازی، امنیت اطلاعات و مصرف انرژی بررسی شده است. مدل برنامه‌ریزی خطی نهایی با استفاده از الگوریتم‌های فراابتکاری نسخهٔ دوم الگوریتم ژنتیک، با مرتب‌سازی نامغلوب (NSGA-II) و نسخهٔ دوم، مبتنی بر قوت پارتو (SPEA-II) بهینه‌سازی

[1] Big Data

شده است. نتایج حل مدل با استفاده از الگوریتم‌های NSGA-II وSPEA-II ، سود زنجیرهٔ تأمین مجازی را به‌ترتیب ۱۰^۶ * 93/9و ۱۰^۶ * 23/4و سرعت پردازش اطلاعات را به‌ترتیب ۳۳۷/۴۸ و ۹۴/۰۷ واحد نشان داد. به این ترتیب، الگوریتم NSGA-II در سودسازی زنجیرهٔ تأمین عملکرد بهتری دارد[1].

(احمدرضا توسلی،۱۴۰۳ )در خصوص تشخیص اخبار جعلی با اخبار واقعی از روش های یادگیری ماشین با نظارت استفاده نموده است و این روش به صورت همزمان با توسعه اینترنت، ظهور و پذیرش گسترده مفهوم رسانه‌های اجتماعی، نحوه شکل‌گیری و انتشار اخبار را تغییر داده است. اخبار سریع‌تر، کم‌هزینه‌تر و به راحتی با رسانه‌های اجتماعی قابل دسترسی هستند. این تغییر با معایبی نیز همراه بوده است. به ویژه، محتوای فریبنده، مانند اخبار جعلی ساخته شده توسط کاربران رسانه‌های اجتماعی، به‌طور فزاینده‌ای خطرناک می‌شود. هدف اصلی این مقاله ارائه راه‌حلی برای چالش تشخیص تفاوت بین اخبار واقعی و جعلی است .این مقاله از نظر هدف کاربردی است و از نظر شیوه اجرا، گردآوری و تحلیل داده‌ها پژوهشی کمی است، که با انتخاب دو مجموعه داده مجموعه داده منبع باز از وب‌گاه کگل در سال ۲۰۲۴ (به‌عنوان جامعه آماری) و پیش‌پردازش این مجموعه داده با استفاده از سامانه رپیدماینر و بکارگیری ماتریس درهم‌ریختگی مبتنی بر پنج مدل یادگیری ماشین نظارت شده، مدل سازی و اجرا شده است. در این مقاله پس از تجزیه و تحلیل داده‌ها و محاسبه معیارهای ارزیابی صحت، یادآوری، دقت و امتیاز (F1میانگین همساز) و مقایسه نتایج مشخص گردید که مدل یادگیری ماشین جنگل تصادفی در معیارهای صحت و دقت به ترتیب با ۹۸.۳ درصد و ۹۷.۹ درصد و مدل یادگیری تقویت گرادیان در محاسبه معیارهای امتیاز F1و یادآوری به ترتیب با ۹۷.۷ درصد و ۹۸.۷ درصد بهترین نتیجه در تشخیص اخبار جعلی دارند[2].

(پرستو اسمعیل زاده ملاباشی و محسن عبدالهی،۱۳۹۹) در مقاله خود برای جلوگری حملات سایبری و نقص اصل عدم مداخله را مورد بررسی قرار داده و پیشرفت فناوری موجب مواجهۀ روزافزون دولت‌ها با حملات سایبری شده است .بیشترین حملات سایبری که دولت‌ها با آن مواجه‌اند، از نوع حملات سایبری نفی یا محروم‌سازی از سرویس توزیع‌شدۀ اینترنتی است. این‌گونه حملات آثار مخرب مستقیم و آنی ندارند، به همین دلیل ارزیابی آنها در قالب ممنوعیت توسل به زور و حملات مسلحانه قرار نمی‌گیرد و معمولاً دولت‌ها نیز با توجه به شدت کمتر آنها در برخی موارد حتی از پیگیری و شناسایی عاملان حملات صرف‌نظر می‌کنند. با اینکه قواعد مستقیم و صریحی در مورد حملات سایبری و نظم بخشیدن به آنها وجود ندارد، نظر به تبعات چنین حملاتی حتی با شدت کم و اقتضای ارزیابی حقوقی این حملات، با بررسی مقررات فعلی حقوق بین‌الملل به این نتیجه می‌رسیم که بعضی از این‌گونه حملات غیرمخرب را می توان با اصل ممنوعیت مداخله به نظم درآورد و در صورت احراز عاملان و انتساب آن حملات به دولت، مسئولیت بین‌المللی دولت‌ها را در مراجع بین‌المللی مطرح کرد .به‌عبارت دیگر، در مقالۀ حاضر سعی بر آن است تا نشان داده شود که صرفاً حملات سایبری شدید ناقص مقررات حقوق بین‌الملل حاضر نیستند[3].

(Zahedi,2024 )در مقاله خود ابزار هوش مصنوعی را برای حفاظت اطلاعات در اینترنت مورد بررسی قرار داده است و در این پژوهش هوش مصنوعی شبیه‌سازی هوش انسانی به وسیله سیستم‌ها و دستگاه‌های کامپیوتری تعریف شده است. در دهه گذشته این فناوری با سرعتی شتابان توسعه یافته و از این طریق توانسته بسیاری از وظایف انسان‌ها را به خوبی انجام دهد. این فناوری با تقلید از شبکه عصبی مغز انسان توانسته به بسیاری از پرسش‌های روزمره پاسخ دهد و روز به روز کاربردهای بیشتری بیابد. برخی از مزایا و خدمات هوش مصنوعی عبارت‌اند از: توانایی یادگیری، تجزیه و تحلیل و تفسیر داده‌ها با دقت و سرعت بالا، کمک به تصمیم‌گیری و ارتقاء کارآمدی و بهره‌وری، صرفه‌جویی در زمان، ایفاء نقش مشاور در زمینه‌های مدیریتی، مالی، پزشکی، امنیتی و غیره البته هوش مصنوعی دارای خطرات و تهدیدهایی نیز هست که صاحب‌نظران در این زمینه مکرراً هشدارهایی داده‌اند. نگرانی‌هایی از جمله استفاده نامناسب از داده‌های شخصی و اطلاعات بیومتریک افراد و تجاوز به حریم خصوصی آنها، دور زدن قوانین و مقررات موجود، افزایش امکان تقلب در رقابت‌ها، تهدید جان افراد از طریق تولید سلاح‌های هوشمند خودمختار و خودران، انجام حملات سایبری و سایر تهدیدات امنیتی وغیره برای استفاده مؤثر از این فناوری، لازم است که از یک سو، حاکمیت و قوای سه‌گانه کشور، مسئولانه و مدبرانه به وظایف خود عمل کنند و زمینه مناسب برای بهره‌مندی آحاد جامعه از آن را فراهم نمایند و از سوی دیگر سازمان‌های جهانی نیز متوجه ابعاد منفی این فناوری بوده و ضرورت رعایت الزامات اخلاقی در استفاده از این پدیده چندوجهی و پرکاربرد را مورد تأکید قرار دهند[4].

(khodad halili,2022) در مقاله خود در مورد عصر فناوری اطلاعات و ارتباطات، توسعه روزافزون فناوری های نوین سایبری، همه عرصه های زندگی بشر از جمله دفاعی و نظامی را به طور اساسی تغییر داده است .شناخت چالش‌ها و تهدیدات این فناوری‌های نوظهور و آماده‌سازی برای رویارویی هوشمند با غول‌های فناوری، دغدغه همیشگی سیاست‌گذاران و ذینفعان سایبری است .بنابراین آینده پژوهی، مطالعه اکتشافی عمیق و بررسی و تحلیل اسناد و گزارش های معتبر جهانی در این زمینه ضروری است .در این پژوهش پس از بررسی فناوری‌های اینترنت اشیا، رایانش ابری، کلان داده و سیستم‌های سایبری فیزیکی، نمونه‌هایی از کاربردهای آن‌ها در سازمان‌های نظامی و دفاعی و همچنین تهدیدات سایبری مورد توجه قرار گرفته است .این تحقیق از نظر هدف کاربردی-توسعه ای و از نظر روش گردآوری داده ها توصیفی- موردی- خاص است .برای جمع آوری داده ها از منابع کتابخانه ای، سایت های علمی معتبر، اسناد و گزارش های داخلی و خارجی استفاده شده است و رویکرد مورد استفاده برای تحلیل داده ها کیفی است .نتایج این پژوهش نشان می دهد که اگرچه استفاده از این فناوری ها می تواند کیفیت و کارایی محیط های عملیاتی سازمان های پدافندی را بهبود بخشد .اما استفاده صحیح از مزایا و قابلیت های آنها به دلیل تهدیدات سایبری نیازمند زیرساخت های ارتباطی و شبکه های اختصاصی است .بنابراین، برای جلوگیری از سهل انگاری استراتژیک در پذیرش و استفاده از این فناوری ها، باید نگرانی ها در مورد دسترسی به اطلاعات حساس، تضمین امنیت داده ها و عملکرد جذاب و وسوسه انگیز را با هم تطبیق داد[5].

**روش تحقیق**

در این پژوهش با توجه به حجم بیشمار اطلاعات در فضای اینترنت و نرم افزار های کاربردی که افراد در طول روز سرکار دارند و ازطرفی افرادی سودجو نیز در تلاش هستند به این سیستم های کامپیوتری نفوذکنند و اطلاعات افراد را فاش و سرقت کنند. به همین دلیل به جهت تشخیص حملات و جلوگیری از نفوذ از ابزارهای هوش مصنوعی برای امنیت اطلاعات استفاده می کنیم به همین دلیل با آموزش هوش مصنوعی روی ایمیل های فیشینگ، فایل های مخرب و رفتارهای خطرناک اپلیکشن ها می توانیم میزان قابل قبولی از شناسایی و تشخیص حملات برسیم، اما مشکل اصلی در حوزه هوش مصنوعی این است که این حوزه به شدت پویا است و مهاجمین مدام در حال ساخت متدهای کلاهبردارانه جدید هستند باید این مدل را برای اثرگذاری و کارایی بیشتر از نو تربیت شوند نیازمند مجموعه داده برچسب گذاری شده از مجموعه ای از داده های تازه و تایید شده رفتار مخرب هستیم که با تربیت هوش مصنوعی روی سرورهای نرمال و فعالیت ایستگاه کاری می توانیم انحرافات نرم و عرف را شناسایی کنیم.

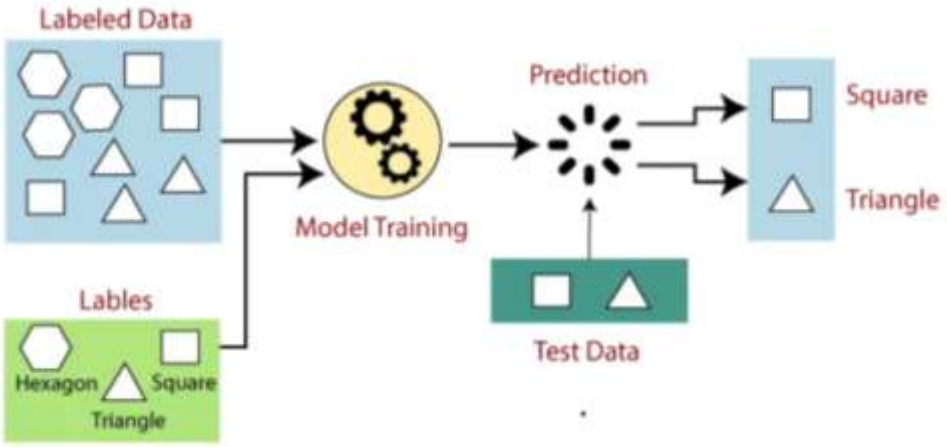
در این روش پیشنهادی هوش مصنوعی به عنوان شریک تحلیل گر بیان می کنیم این ابزار هوشمند نمی تواند تماماً مسئول جستجو تهدیدهای سایبری دانست اما می توانیم حجم کار انسان را با تحلیل مستقل هشدارهای ساده SIEM و کمک تحلیل گران در سایر موارد کم کنیم که با فیلتر کردن موارد ثبت کاذب و آموزش هشدارهای SIEM و احکام تحلیل گران می تواند FP ها را کامل قابل اعتماد فیلتر می کنیم و اولویت بندی هشدارها همان موتور ماشین یادگیری فقط FP را فیلتر نمی کند بلکه احتمال اینکه یک رویداد شناسایی شده نشان دهنده فعالیت مخرب جدی باشد را ارزیابی می کند سپس هشدارهای حیاتی برای تجزیه وتحلیل اولویت بندی شده به کارشناسان ارسال می شود و از طرف دیگر احتمال تهدید را می تواند به عنوان یک شاخص بصیری نشان می دهد و به تحلیل گر کمک می کند تا مهمترین هشدارها را اولویت بندی می کند.

در بحث شناسایی ناهنجاری هوش مصنوعی می تواند به سرعت در خصوص ناهنجاری های زیرساخت محافظت شده با ردیابی پدیده هایی همچون تورم هشدارها، کاهش یا افزایش شدید جریان تله متری از حسگرهای تعیین شده یا تغییراتی در ساختار آن هشدار می دهد و برای شناسایی رفتار مشکوک جستجو ناهنجاری های دلخواه در شبکه مشکلات قابل توجهی دارد برخی از سناریوها به اتوماسیون کمک می کند. یادگیری ماشین از قوانین ایستا بهتر عمل می کند و شناسایی استفاده غیرقانونی حساب از زیرمجموعه های غیر معمول شناسایی و دسترسی ناهنجار به فایل های سرورها و اسکن ها و نیز جستجو برای حملات می شود.

مدل های زبانی بزرگ ترند حوزه هوش مصنوعی هستند و همچنین توسط سازمان های امنیت اطلاعات نیز به طور گسترده تست شده اند تلاش های مجرمانه مانند تولید ایمیل های فیشینگ و بدافزار با استفاده از چت جی بی تی[۲] به این آزمایش های جالب و فراوان در استفاده از این مدل های زبانی بزرگ توجه می کند این آزمایش ها در این حوزه شامل ایجاد توضیحات دقیق درباره تهدید های سایبری، تهیه پیش نویس گزارش های بررسی رخداد، جستجوی فازی در آرشیو داده ها و سیاهه های مربوط از طریق چت، تولید تست و کد برای fuzzing، تحلیل اولیه کد منبع دیکامپایل شده در مهندسی معکوس، رفع ابهام و توضیح خطوط فرمان طولانی( سرویس MDR ما در حال حاضر از این فناوری استفاده می شود و ایجاد نکات و هشدارهایی برای نوشتن قوانین تشخیص و اسکریپ[۳] می باشد.

در این روش پیشنهادی روی عملکرد ارزیابی مانور داده شده است و نتایج نشان می دهد چنین راهکارهایی به تدریج و در مراحل مختلف پیاده سازی آن با ارزیابی مقدماتی پتانسیل پس اندازها و ارزیابی بر جزئیات سرمایه گذاری در زمان و کیفیت نتیجه می شوند.

بعد از جمع آوری داده های کافی در جهت تشخیص حملات پیچیده در اتوماسیون های سازمانی میلیون ها رخداد ها از حسگرهای کل شبکه انجام می شود و بعد از طبقه بندی آن و فیلتر کردن و تمیز دادن داده ها با استقاده از الگوریتم SIEM در مورد فعالیت های بالقوه مخرب را تقطیر می شوند. با توجه به مجموعه داده های Kaspersky MDR به زیرساخت های client ها رخداد روزانه را تولید می کند. برای فیلتر و تمیز دادن داده ها این تحلیل گر خودکار توسعه داده می شود، این سیستم یادگیری ماشین با نظارت روی هشدار سیستم SIEM ترکیب شده و با این تحلیل خودکار اتوماسیون روی هر رخداد تربیت شده است هدف از این آموزش هوش مصنوعی با اطمینان مثبت کاذب های تولید شده توسط فعالیت های قانونی شبکه شناسایی می کند. این حوزه نسبت به شناسایی تهدید از پویایی کمتری برخوردار است راحت تر می شود این یادگیری ماشین مبتنی بر catboost است یک کتابخانه تقویت کننده گرادیان، تحلیل خودکار آموزش دیده هشدار ها را فیلتر می کند وفقط آن نودهایی که در شبکه احتمال وقوع رخداد واقعی بالاتر از حد آستانه دارند را تعیین می کند و میزان خطای قابل قبولی را برای بررسی انسان است ارسال می کند در نتیجه حدود ۳۰ درصد از هشدارها توسط تحلیل گر خودکار مدیریت می شود و توسط تیم اتوماسیون سازمانی برای کارهای پیچیده تر آزاد می کند. عملکرد این یادگیری در شکل ۱ نشان می دهد.




**شکل ۱ - عملکرد الگرویتم یادگیری بانظارت در روش پیشنهادی**

در شکل ۱ عملکرد این الگوریتم به شکلی است که ابتدا داده ها را جمع آوری می کند سپس از طریق الگوریتمSIEM داده ها را پاکسازی می کند پس از آن ۷۵ درصد داده ها برای آموزش و ۲۵ درصد داده ها برای تست آماده می شود. ویژگی های ورودی مجموعه داده آموزشی

[2] Chat gpt
[3] script

را تعیین کنید، که باید دانش کافی داشته باشد تا مدل بتواند خروجی را از طریق آن ها به طور دقیق پیش بینی کند و با استفاده از مجموعه داده تست دقت و صحت مدل مورد ارزیابی می شود.

**یافته ها**

در این بخش، ارزیابی روش پیشنهادی با استفاده از مجموعه داده های Kaspersky daily به صورتی می باشد که خوشه های Big Dataدر سرورهای sql، معمولا وظیفه های هوش مصنوعی و یادگیری ماشینی را بر داده‌های ذخیره شده در HDFS Storage Poolها وPool های مختص به داده‌ها، فعال می‌کنند و کاربران می‌تواند با استفاده از نرم افزار های R ، Python، Scala یا جاوا از Sparkو ابزار هوش مصنوعی ساخته شده در سرور SQL بهره می برند با توجه به مجموعه داده های آموزشی تمام داده هایی که هرزنامه می باشد را طبقه بندی می کند و یک نمونه جدید ایجاد می کند و با آن نمونه جدید تست را جهت بدست آوردن دقت و صحت کار ایجاد می کند و در شکل ۲ روند طبقه بندی این هرزنامه را نشان می دهد.


شکل ۲- طبقه بندی هرزنامه با توجه به مجموعه داده آموزش

با توجه به شکل ۲ و براساس این طبقه بندی در طول آموزش به دنبال الگوهایی می باشد که در داده‌هایی که با خروجی‌های مورد نظر ارتباط دارند کار را به سمت جلو حرکت می دهد و پس از آموزش، یک الگوریتم یادگیری تحت نظارت ورودی‌های جدید نادیده گرفته خواهد شد و تعیین خواهد کرد که کدام ورودی‌های جدید براساس داده‌های آموزشی قبلی طبقه‌بندی خواهند شد. هدف از یک مدل یادگیری تحت نظارت، پیش‌بینی برچسب صحیح برای داده‌های ورودی جدید است.

**بحث و نتیجه‌گیری**

در این روش پیشنهادی می توانیم در برخی از سیستم ها براساس رفتار کاربران ، نودهای مشکوک را شناسایی کنیم و با این روش ریسک ورود به شبکه را ارزیابی می کنیم و اقدامات احتیاطی بیشتر با احراز هویت قوی از جمله امضای دیجیتال اعمال کنیم. با توجه به این که امروزه بیشترین حمله در فضای اینترنت حمله فیشینگ است با این روش پیشنهادی می توانیم راه حلی برای افزایش امنیت با وظایف تجاری و ریسک تجاری بیان کنیم اما این تکنیک های استفاده شده در این روش به طور مداوم در حال تکامل است و مهاجمان در فکر راهی برای حمله به سیستم های کامپیوتری می باشد و باید سازمان ها و شرکت ها با آخرین تهدیدات و اجرای استراتژی دفاعی چند لایه سیستم های کامپیوتری خود را محافظت کنند و برای اینکار مستلزم هوشیاری، آموزش و ابزار مناسب است که با اولویت بندی تمام حوزه ها در سازمان ها می توانیم از خطر حمله سایبری موفق به میزان قابل توجهی کاهش دهیم و اعتماد کاربران و ذینفعان خود را حفظ کنیم.

## منابع

شامبیاتی, هانیه, شفیعی نیک آبادی, محسن, خاتمی فیروز آبادی, سیدمحمدعلی ,رحمانی منش, محمد & ,صابری, سارا. (۱۴۰۱). مدلی جهت بهینه‌سازی عملکرد پردازش اطلاعات در زنجیرهٔ تأمین مجازی، مبتنی بر اینترنت اشیا .*پژوهش در مدیریت تولید و عملیات* :doi .1-24 ,(1)13 , 10.22108/jpom.2022.129445.1385.

ترسلی. (۲۰۲۴). کاربرد هوش مصنوعی در تشخیص اخبار جعلی. فصلنامه آماد و فناوری دفاعی, ۷(۲), ۸۵-۱۱۲.

اسمعیل‌زاده ملاباشی, پرستو & ,عبدالهی, محسن. (۱۳۹۹). حملات سایبری و نقض اصل عدم مداخله .*فصلنامه مطالعات حقوق عمومی دانشگاه تهران* ,(2)50 , 711-736. doi: 10.22059/jplsq.2019.230048.1499

Zahedi, S. (2024). Management with AI and on AI. *Development of humanities*, *4*(8), 45-58. doi: 10.22047/hsd.2024.194458.

Halili, K. (2022). Emerging cyber technologies and threats posed by their use in military-defense organizations. *War Studies*, *3*(11), 97-121.

# Investigating big data in detecting attacks on the Internet using artificial intelligence

Mohmmadreza Hosseinzadeh Moghaddam

Ph.D. student, software engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran

## Abstract

Today, artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. AI systems generally work by taking large amounts of labeled training data, analyzing the data for correlations and patterns, and using these patterns to predict future situations. They can learn to create real interactions with people, or a diagnostic tool. Image can learn to identify and describe objects in images by reviewing millions of examples. Artificial intelligence enables machines to adapt to new inputs and perform human-like tasks. Most examples of artificial intelligence you hear about today rely heavily on deep learning and natural language processing. Using these technologies, computers can be trained to perform specific tasks by processing large amounts of data and recognizing patterns in the data. In this method, we use supervised machine learning because, in supervised learning, the machine learns by using labeled data and having correct answers. In this method, we use the SIEM algorithm to filter data and use it in the organization space. It can perform well in detecting phishing attacks.

Keywords: SIEM Algorithm, Machine Learning with Supervised, Artificial Intelligence, Bigdata and Internet