تشخیص نفوذ در شبکه با استفاده از تکنیک‌های یادگیری ماشین

حسین فقیه علی آبادی [۱]، علی زهدی [۲]

[۱] کارشناسی ارشد شبکه‌های کامپیوتری، دانشگاه ارومیه، hosseinfaghih1995@gmail.com
[۲] کارشناسی ارشد شبکه‌های کامپیوتری، خواجه نصیر الدین طوسی، mralizohdi@gmail.com

چکیده

با افزایش استفاده از منابع اینترنتی، مهاجمان سایبری از راه‌های جدیدی برای حمله به خدمات شبکه استفاده می‌کنند، بنابراین امنیت شبکه در حال تبدیل شدن به بخشی اجتناب ناپذیر از سیستم شبکه است. برای شناسایی حملات، به سیستم تشخیص نفوذ قوی نیاز است. سیستم تشخیص نفوذ ابزاری است که هر بسته را عمیقاً تجزیه و تحلیل می‌کند تا با نظارت بر یک شبکه یا یک سیستم، فعالیت‌های مخرب را شناسایی کند. از شش الگوریتم یادگیری ماشین برای به کارگیری سیستم تشخیص نفوذ استفاده می‌شود. در این پژوهش از روش‌های یادگیری ماشین به برای طبقه‌بندی باینری مجموعه داده UNSW-NB15 و طبقه بندی چندکلاسه مجموعه داده MQTT استفاده شده است. خروجی الگوریتم‌های پیشنهادی با استفاده از چهار معیار ارزیابی دقت، صحت، فراخوانی و امتیاز F1 ارزیابی می‌شود. الگوریتم XGBOOTS بهترین عملکرد در UNSW-NB15 و الگوریتم XG و RF بهترین عملکرد در MQTT نمایش داده است.

واژه های کلیدی

امنیت، شبکه، سیستم تشخیص نفوذ، هوش مصنوعی، یادگیری ماشین

۱.مقدمه

احتمال سرقت اطلاعات، افشا، وقفه، عدم حفظ یکپارچگی و غیره، از طریق رشد و فراگیر شدن اینترنت گسترش یافته است. به دلیل حجم عظیم داده در شبکه سیستم در برابر انواع حملات آسیب پذیر می‌شود و نفوذهای مختلف روز به روز در حال افزایش می‌باشند. تشخیص نفوذ یک گام اساسی برای جلوگیری از نفوذ یا سوء استفاده از داده‌های شبکه است. برای دفاع در برابر نفوذهای متعدد شبکه و فعالیت‌های مخرب، روش های زیادی توسعه داده شده است. تشخیص نفوذ شبکه یکی از امیدوارکننده‌ترین روش‌ها برای محافظت از شبکه در برابر رفتارهای نفوذ در نظر گرفته می‌شود[۱]. نفوذ یک فعالیت مخرب است که هدف آن به خطر انداختن محرمانه بودن، یکپارچگی یا در دسترس بودن هر یک از اجزای شبکه به عنوان تلاشی برای برهم زدن سیاست امنیتی شبکه است [۲]. سیستم تشخیص نفوذ با طبقه‌بندی داده‌ها در دسته‌های مختلف، رفتارهای نفوذی را از فعالیت‌های عادی شبکه متمایز می‌کند. بر اساس استراتژی های تجزیه و تحلیل و روش های تشخیص،[1]IDS ها به سیستم‌های تشخیص نفوذ سوءاستفاده و سیستم های تشخیص نفوذ ناهنجار دسته‌بندی می‌شوند. بر اساس منبع داده، IDS ها به سیستم های تشخیص مبتنی بر شبکه و مبتنی بر میزبان طبقه بندی می‌شوند[۳]. چندین روش یادگیری ماشین برای توسعه تشخیص نفوذ موثر و هوشمند پیشنهاد شده است. با این حال، مطالعات کمی وجود دارد که رویکردهای ML را برای تشخیص نفوذ با دقت بالا ارزیابی کنند. به طور معمول، الگوریتم‌های ML را می‌توان به سه دسته تحت نظارت، بدون نظارت و نیمه نظارت طبقه بندی کرد. به طور خاص، طبقه‌بندی‌کننده‌های ML معمولاً به عنوان یک جعبه سیاه استفاده می‌شوند، جایی که نتایج گزارش‌شده ممکن است در نتیجه تطبیق بیش از حد مدل برای یک مجموعه داده خاص به‌دست آیند، بنابراین نتایج چندان قابل تعمیم و تکرار نیستند[۴]. سهم اصلی ما در این مقاله ارائه یک ارزیابی تجربی دقیق از الگوریتم‌های یادگیری ماشین در زمینه تشخیص نفوذ است. پس از مقدمه‌ای که در این بخش بیان کردیم، در قسمت دوم کارهای گذشته را بررسی می‌کنیم، در قسمت سوم روش پیشنهادی این مقاله مطرح می شود. در قسمت چهارم به ارزیابی روش پیشنهادی و در قسمت آخر به نتیجه گیری حاصل از این پژوهش می‌پردازیم.

۲. بررسی کارهای گذشته

در [۵] هدف دستگاه IoT در شبکه هوشمند را مورد بحث قرار دادند که می‌تواند بسیار آسیب پذیر باشد و حتی مهاجمان می‌توانند داده‌های حسگرها را تغییر دهند. حملات عمده‌ای که در دستگاه های اینترنت اشیا رخ می‌دهد عبارتند از حمله فیزیکی، حمله کانال جانبی، حمله محیطی، حملات رمزنگاری، حمله سیاه چاله، حمله Sybil و

[1] Intrusion Detection System

غیره. در [۶] به امنیت اینترنت اشیا در شبکه‌های 5G اختصاص دارد که با تعداد زیادی دستگاه و نرخ بالای انتقال داده مشخص می شود. طرح ها از منظر به کارگیری ML در نظر گرفته می شوند. فن‌آوری‌های ML، مانند فرآیند تصادفی خودکار، پیش‌بینی فیلتر کالمن، یادگیری تقویتی (مخصوصاً یادگیری Q و Dyna-Q)، طبقه‌بندی کننده AdaBoost و ماشین هسته، مورد بررسی قرار می‌گیرند. در [۷] یک IDS بر اساس تکنیک یادگیری عمیق با استفاده از مجموعه داده NSL-KDD برای تشخیص نفوذ در شبکه طراحی کردند. مدل یاد می‌گیرد و همچنین توانایی تطبیقی برای یافتن الگوهای جدیدی را دارد که قبلاً تفسیر نشده اند. در [۸] از مجموعه داده NSL-KDD و MLP به عنوان مدل ML برای یک سیستم IDS فعال با FL استفاده می‌کند. این رویکرد مبتنی بر مفهوم یادگیری تقلید است که در آن یک مدل دانش آموز با یک مجموعه داده عمومی آموزش می‌بینند که با یک مدل اصلی آموزش دیده با داده‌های حساس برچسب گذاری شده است. در [۹] از مجموعه داده NB-IOT استفاده می شود، که یک رویکرد طبقه بندی دودویی را بر اساس یادگیری نظارت شده و یادگیری بدون نظارت پیشنهاد می‌کند. در [۱۰] به شدت چنین حملاتی را مورد توجه قرار داده و استراتژی های بهینه را برای انجام چنین اقدامات بدخواهانه در محیط‌های IoT ارائه می‌کند. چارچوب امنیتی طراحی شده بر روی مجموعه داده NSL-KDD آموزش داده شده است. علاوه بر این، برخی از الگوریتم‌های آموزشی قابل توجه به نام‌های شبکه چند جمله‌ای عمیق پشته‌ای و بهینه سازی عنکبوتی در نظر گرفته شده‌اند. چارچوب پیشنهادی با حملات ذکر شده با دقت تشخیص تقریباً ٪۹۶,۰۲، ٪۹۶,۳ مواجه می‌شود.

۳. روش پیشنهادی

یادگیری ماشین زیرمجموعه‌ای از هوش مصنوعی است. ML باعث می‌شود سیستم بدون برنامه‌ریزی صریح، توانایی‌های خودکار خود را از تجربه یاد بگیرد و بهبود بخشد. برای سیستم تشخیص نفوذ، الگوریتم ML در تشخیص حملات برای حجم عظیمی از داده‌ها در زمان کمتری با دقت بیشتری کار می‌کند. در این بخش کار پیشنهادی را مورد بحث قرار می‌دهیم که در آن از چهار طبقه‌بندی کننده یادگیری ماشین برای علامت‌گذاری بسته‌ها در حالت باینری با استفاده از مجموعه داده UNSW– NB15 و انجام شده است. همچنین از شش طبقه‌بندی کننده برای علامت‌گذاری بسته‌ها در حالت چند کلاسه با استفاده از مجموعه داده MQTT-IDS انجام شده است.

۱,۳- الگوریتم پیشنهادی

چالش‌های اولیه‌ای که IDS با آن مواجه است نرخ بالای هشدار کاذب (منفی کاذب و مثبت کاذب) و عدم پاسخگویی در زمان واقعی است. الگوریتم‌های یادگیری ماشین قدرت مقابله با چنین چالش‌هایی را دارند. از تکنیک‌های یادگیری ماشین می‌توان برای ساخت IDS هوشمند استفاده کرد زیرا می‌تواند حملات شناخته شده و ناشناخته را با سرعت بالا، حداکثر دقت و حداقل نرخ هشدار نادرست شناسایی کند، بنابراین الگوریتم‌های یادگیری ماشین می‌توانند برای تقویت IDS با قابلیت‌های پیشرفته استفاده شوند.

رگرسیون لجستیک یک الگوریتم طبقه بندی ML نظارت شده است که برای مشاهده مجموعه گسسته کلاس‌ها استفاده می شود. این الگوریتم تحلیل پیش‌بینی بر اساس مفهوم احتمال است. تابع لجستیک از تابع هزینه استفاده می‌کند که تابع سیگموئید نامیده می شود. از تابع Sigmoid برای ترسیم مقادیر پیش بینی شده به احتمالات بین ۰ و ۱ استفاده می‌کند. معادله (۱) به صورت تابع زیر تعریف می‌شود:

(۱)

$$sigmoid\ (x) = \frac{1}{1+ e^{x}}$$

در جایی که Sigmoid(x) خروجی بین ۰ و ۱ است، x ورودی تابع و e پایه ورود به سیستم طبیعی است.

جنگل تصادفی یک الگوریتم نظارت شده غیرخطی پیچیده است که برای طبقه بندی و رگرسیون استفاده می شود. این درخت تصمیم-گیری بسیاری را در آموزش مدل ایجاد می‌کند و نتایج پیش بینی‌ها از همه درختان برای ایجاد یک نتیجه جمع می‌شود، بنابراین به عنوان تکنیک‌های Ensemble ذکر می‌شود. طبقه‌بندی کننده‌های RF به این صورت عمل می‌کنند که هر چه تعداد درخت‌ها در مدل بیشتر باشد، دقت بالاتری را به همراه خواهد داشت و مدل بیش از حد برازنده نمی‌شود. دقیقاً مانند درخت تصمیم برای ساختن درخت عمل می کند با این تفاوت که درخت تصمیم با استفاده از کل مجموعه داده با در نظر گرفتن همه ویژگی‌ها ساخته می‌شود، در حالی که یک جنگل تصادفی به طور تصادفی مشاهدات و ویژگی‌های خاص را برای ساخت درخت-های تصمیم گیری متعدد انتخاب می‌کند و سپس آن‌ها را با هم ادغام می‌کند تا دقت و پیش بینی پایداری بالاتری به دست آورد.

K–نزدیکترین همسایه اصل این مدل طبقه‌بندی یک نقطه بر اساس فاصله آن از k نزدیکترین نقطه همسایه است. این تصمیم با اکثریت آرای همسایگان گرفته می‌شود. برای طبقه‌بندی در مرحله اول، هر قطعه داده را به عنوان یک گره در فضای n بعد قرار می‌دهد، جایی که n تعداد ویژگی‌های یک مجموعه داده است. در مرحله دوم، محاسبه فاصله اقلیدسی، بین ورودی داده و هر گره موجود نشان داده شده است. در مرحله سوم، مرتب سازی داده‌ها به ترتیب صعودی و محاسبه اکثریت

برچسب‌های k نزدیکترین همسایه الگوریتم مرتب سازی پیچیدگی محاسبات را تعیین می‌کند.

XGBoost اخیراً در حوزه یادگیری ماشین بکار گرفته می‌شود. الگوریتم XGBoost یک پیاده سازی از تقویت گرادیان درخت تصمیم-گیری است که برای سرعت و کارایی بالا طراحی شده است. ویژگی‌های XGBoostرا می‌توان به ویژگی‌های مدل، ویژگی‌های سیستم و ویژگی‌های الگوریتم اشاره کرد. XGBoosting در مقایسه با پیاده سازی تقویت گرادیان، بسیار سریع است. می‌توان XGBoost را با سایر پیاده سازی‌های تقویت گرادیان و درختان تصمیم‌گیری مقایسه کرد. XGBoost توان پیش‌بینی بسیار بالایی دارد که آن را تبدیل به بهترین گزینه برای دقت در رویدادهای مختلف می‌کند چرا که شامل مدل خطی و الگوریتم یادگیری درختی است. این الگوریتم تقریباً ۱۰ برابر سریع‌تر از الگوریتم‌های موجود ارتقای گرادیان است. این الگوریتم شامل تابع‌های عینی مختلف، رگرسیون، کلاس‌بندی و رتبه‌بندی است. این الگوریتم به کاهش مدل‌های بزرگ کمک می‌کند.

ماشین بردار پشتیبان یکی از روش‌های یادگیری بانظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. مبنای کاری دسته‌بندی کننده SVM دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کنیم ابرصفحه ای را انتخاب کنیم که حاشیه اطمینان بیشتری داشته باشد.

فرآیندهای گاوسی یک الگوریتم یادگیری ماشینی طبقه‌بندی است. فرآیندهای گاوسی تعمیم توزیع احتمال گاوسی است و می‌تواند به عنوان پایه ای برای الگوریتم‌های پیچیده یادگیری ماشین غیر پارامتری برای طبقه بندی و رگرسیون استفاده شود.

۲,۳- مجموعه داده

اکثر مدل‌های IDS فقط برای مجموعه داده‌های خاص مناسب هستند در حالی که برای مجموعه داده دیگری ضعیف عمل می‌کنند. از این رو در این‌کار، دو مجموعه‌داده عمومی تشخیص نفوذ برای آزمایش استفاده خواهد شد.

مجموعه داده UNSW-NB15: مجموعه داده UNSW – NB15 روش ترکیبی شامل رفتارهای نرمال و حمله یک ترافیک شبکه با استفاده از ابزار IXIA Perfect Storm تولید کرده است. این داده از دو سرور در ابزار تولید ترافیک IXIA استفاده کرده است که در آن یک سرور فعالیت‌های نرمال را ایجاد می‌کند ودیگری فعالیت‌های مخربی را در شبکه ایجاد می‌کند. مجموعه داده UNSW-NB15 شامل ۴۲ ویژگی همراه با برچسب کلاس 1 , 0 می‌باشد. که برای Normal مقدار صفر و برای Attack مقدار یک در نظر گرفته می‌شود[۱۱].

مجموعه داده MQTT: در این داده پنج سناریو ثبت شده که شامل عملیات عادی و چهار سناریو حمله می‌باشد. مهاجم چهار حمله زیر را انجام می‌دهد و هر کدام به طور مستقل ضبط می شود.

- Aggressive scan (Scan A)
- User Datagram Protocol (UDP) scan (Scan sU)
- Sparta SSH brute-force (Sparta)
- MQTT brute-force attack (MQTT BF)
- Normal

داده‌ها با استفاده از tcpdump به دست می‌آیند و بسته‌ها با ضبط ترافیک اترنت و سپس صادرات به فایل های pcap جمع آوری می-شوند[۱۲].

۳,۳- پیش پردازش

عدم تعادل داده‌ها یک مشکل شناخته شده در یادگیری ماشین است، زمانی رخ می‌دهد که توزیع کلاس‌های مختلف بایاس باشند. در یک مجموعه داده نامتعادل، توزیع طبقات مختلف می‌تواند کمی نامتعادل یا شدیداً نامتعادل باشد. هر مدل یادگیری که بر روی یک مجموعه داده به شدت نامتعادل آموزش داده شود، منجر به عملکرد پیش‌بینی ضعیف در برابر کلاس‌های کوچک می‌شود. پیش پردازش شامل تنظیم و عادی سازی داده‌ها است. در مجموعه داده اصلی، مقادیر انواع مختلفی دارند، برخی ویژگی‌های مبتنی بر کاراکتر هستند که نمی‌توان آن‌ها را با مدل یادگیری عمیق پردازش کرد، که این مقادیر برای پردازش داده‌ها باید به نوع عددی تبدیل شوند. به همین ترتیب، برخی از مقادیر ممکن است به دلیل نمایش متفاوت در محدوده مشخص شده قرار نگیرند. برخی از ویژگی‌ها بیش از حد متفاوت هستند و برای طبقه‌بندی نهایی تشخیص نفوذ، نرمال‌سازی و پردازش ویژگی مناسب نیستند، ما در این بخش با روش‌های نرمال سازی، پردازش عددی و OHE[۲] سعی در کاهش این مشکلات داریم.

۴. نتایج

تمام آزمایش‌ها در Google Colab انجام شد و اثربخشی همه طبقه‌بندی‌کننده‌ها در طبقه‌بندی مجموعه داده MQTT و UNSW-NB15 مورد بررسی قرار گرفت. راهکار پیشنهادی، باید از جنبه‌های مختلف مورد ارزیابی قرار گیرد. در این بخش معیارها و نتایج ارزیابی بیان می‌شود.

۱,۴- شاخص‌های ارزیابی

راهکار پیشنهادی، باید از جنبه‌های مختلف مورد ارزیابی قرار گیرد. هر نمونه یا فردی در واقعیت، متعلق به یکی از کلاس‌های مثبت یا منفی است و از سوی دیگر، از هر الگوریتمی که برای دسته‌بندی داده‌ها

[2] One Hot Encoder

استفاده می‌شود، در نهایت هر نمونه عضو یکی از این دو دسته‌بندی خواهد بود. بنابراین برای هر نمونه داده، یکی از چهار حالتی که در ادامه بیان شده، ممکن است اتفاق بیفتد.

- نمونه عضو دسته مثبت باشد و عضو همین کلاس تشخیص داده شود(مثبت صحیح یا True Positive)
- نمونه عضو کلاس مثبت باشد و عضو کلاس منفی تشخیص داده شود(منفی کاذب یا False Negative)
- نمونه عضو کلاس منفی باشد و عضو همین کلاس تشخیص داده شود (منفی صحیح یا True Negative)
- نمونه عضو کلاس منفی باشد و عضو کلاس مثبت تشخیص داده شود(مثبت کاذب یا False Positive)

برای ارزیابی یک مدل، معیارهای بسیار زیادی موجود است که هر کدام در موارد خاص از اهمیتی بیشتری برخوردار هستند. از آن جا که بخش بسیار زیادی از دادگان ما را ترافیک عادی تشکیل داده است، معیارهای بازخوانی و امتیاز F1 اهمیت بالایی پیدا می‌کنند. معیار امتیاز F1 در مجموعه دادگانی که متعادل نیستند و کلاس‌های آن‌ها از تعداد اعضای برابری تشکیل نشده است، اهمیت زیادی پیدا می‌کند. ما در این مقاله از ۴ معیار ارزیابی استفاده کرده‌ایم که توضیحات آن ها به شرح زیر است:

صحت: این معیار رایج ترین معیار مورد استفاده در ارزیابی روش‌های دسته‌بندی است و به صورت رایج در ارزیابی روش‌های یادگیری ماشین گزارش می‌شود و درصد دادگانی که درست دسته‌بندی شده اند را نشان می دهد. در معادله (۲) روش محاسبه این معیار بیان می‌شود:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+TN} \quad (۲)$$

دقت: این معیار به معنی این است که چه درصد از دادگانی که به عنوان اعضای یک کلاس تشخیص داده شده‌اند در واقع متعلق به آن کلاس بوده‌اند. در مساله ما این معیار نشان دهنده این است که چه درصد از جریان‌هایی که به عنوان حمله تشخیص داده شده‌اند واقعا حمله بوده‌اند. معادله (۳) این معیار را محاسبه می‌کند:

$$Precision = \frac{TP}{TP+FP} \quad (۳)$$

بازخوانی: این معیار نشان دهنده این است که چه درصدی از دادگان یک کلاس، تشخیص داده شده‌اند. این معیار در مساله ما اهمیت بالایی دارد زیرا هدف ما شناسایی همه حملات است. این معیار در معادله (۴) بیان می‌شود.

$$Recall = \frac{TP}{TP+FN} \quad (۴)$$

امتیاز F۱: این معیار میانگین همساز Recall و Precision است. این معیار در معادله (۵) به صورت زیر محاسبه می‌شود:

$$F1 = 2\frac{Pre * Recall}{Pre + Recall} \quad (5)$$

۲,۴- تجزیه تحلیل نتایج

در آزمایش مجموعه داده UNSW-NB15 ابتدا در جدول ۱ ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم رگرسیون لجستیک، در جدول ۲ ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم جنگل تصادفی، در جدول ۳ ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم k-نزدیکترین همسایه و در جدول ۴ ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم XGBoost نمایش می‌دهیم، همچنین در شکل ۱ چهار الگوریتم را باهم مقایسه می‌کنیم.

جدول(۱): ماتریس درهم‌ریختگی رگرسیون لجستیک

Predicted lable / lable True	0	1
0	۱۷۷۸۲	۱۹۲۱۸
1	۴۷۵	۴۴۸۵۷

جدول (۲): ماتریس درهم‌ریختگی RF

Predicted lable / lable True	0	1
0	۱۷۷۷۹	۱۹۲۲۱
1	۷۲	۴۵۲۶۰

جدول (۳): ماتریس درهم‌ریختگی KNN

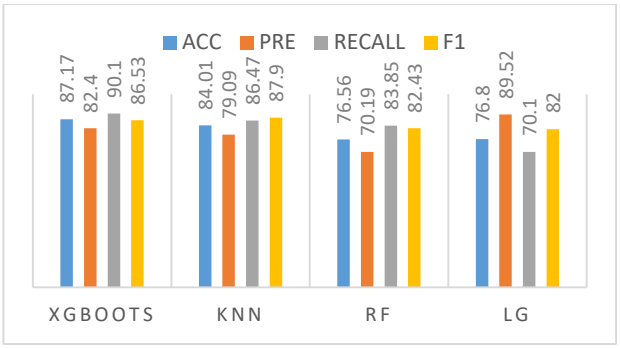
Predicted lable / lable True	0	1
0	۲۵۴۳۸	۱۱۵۶۲
1	۱۵۹۶	۴۳۷۳۶

جدول(۴): ماتریس درهم‌ریختگی XGBOOTS

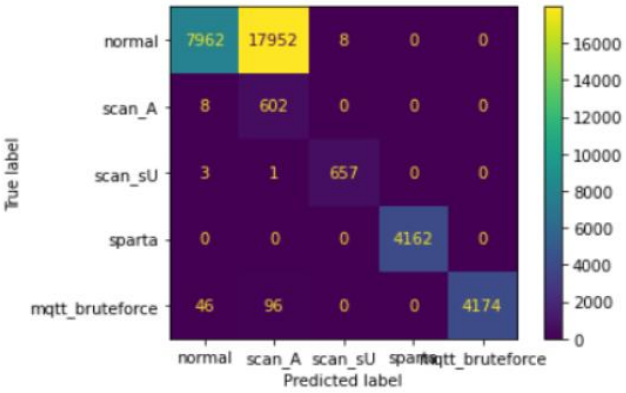
Predicted lable / lable True	0	1
0	۲۷۵۵۸	۹۴۴۲
1	۱۱۱۳	۴۴۲۱۹

در ادامه به نتایج حاصل از آزمایش مجموعه داده MQTT می‌پردازیم. در شکل ۳ ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم Gaussian، در شکل ۴ ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم KNN ، در شکل ۵ ماتریس درهم‌ریختگی حاصل از پیاده

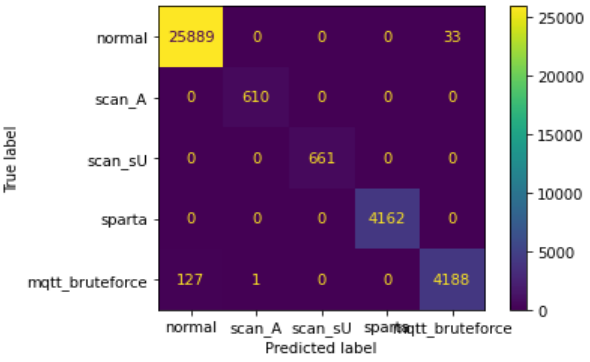
سازی الگوریتم LOG، در شکل ۶ ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتمRF ، در شکل ۷ ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم SVM و در شکل ۸ ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم XG را نشان می‌دهیم. در شکل ۹ و ۱۰ به عنوان نمونه نتایج حاصل از پیاده سازی الگوریتم XG را با پارامترهای ارزیابی نمایش می‌دهیم.



شکل(۱): مقایسه چهار الگوریتم یادگیری ماشین در مجموعه داده UNSW-NB15



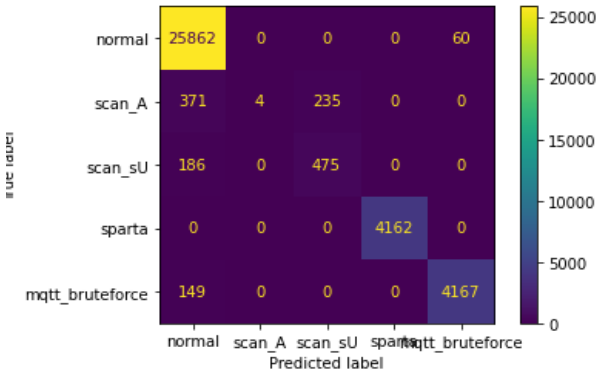
شکل ۳: ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم Gaussian



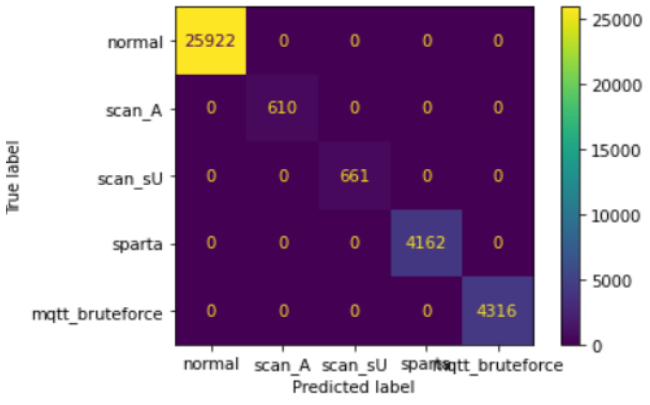
شکل ۴: ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم KNN

```
Accuracy score for normal:          1.0
Accuracy score for scan_A:          1.0
Accuracy score for scan_sU:         1.0
Accuracy score for sparta:          1.0
Accuracy score for mqtt_bruteforce: 1.0
```

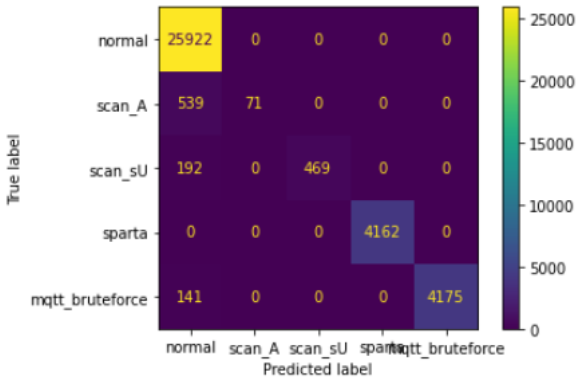
شکل۹: پارامتر دقت برای الگوریتم XG



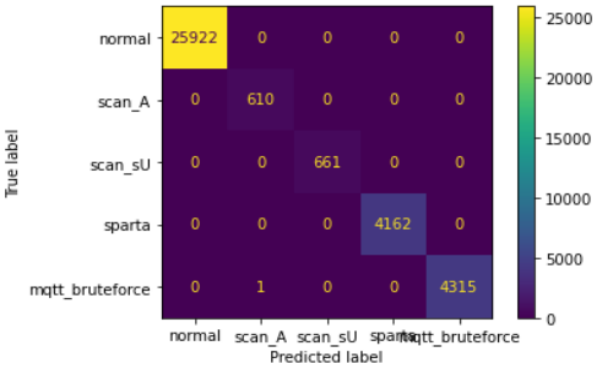
شکل ۵: ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم LOG



شکل ۶: ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم RF



شکل۷: ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم SVM



شکل۸: ماتریس درهم‌ریختگی حاصل از پیاده سازی الگوریتم XG

	precision	recall	f1-score
normal	1.00	1.00	1.00
scan_A	1.00	1.00	1.00
scan_sU	1.00	1.00	1.00
sparta	1.00	1.00	1.00
mqtt_bruteforce	1.00	1.00	1.00
accuracy			1.00
macro avg	1.00	1.00	1.00
weighted avg	1.00	1.00	1.00

شکل۱۰: نمایش سه پارامتر دیگر برای الگوریتم XG

۵. نتیجه گیری

هدف از این پژوهش بررسی چالش‌ها و الزامات مختلف برای ساخت IDS برای مدل‌های IoT بود. در این مقاله، آزمایش‌های تجربی با استفاده از طبقه‌بندی‌کننده یادگیری ماشین به نام‌های رگرسیون لجستیک، جنگل تصادفی، k-نزدیکترین همسایه، ماشین بردار پشتیبان، Gaussian و XGBoots برای برای آزمایش و ارزیابی کارایی و عملکرد انجام شد. در ابتدا مجموعه داده‌های UNSW-NB15 و MQTT برای انتخاب ویژگی‌های مربوطه برای افزایش کارایی و کاهش زمان آموزش پیش پردازش شدند. در مجموعه داده UNSW-NB15 بر اساس نتایج به دست‌آمده در حالت طبقه‌بندی باینری، بیشترین میزان ACC برای الگوریتم XGBoots، بیشترین میزان PRE برای الگوریتم LG، بیشترین میزان RECALL برای الگوریتم XGBoots و بیشترین میزان امتیاز F1 برای الگوریتم KNN ثبت شده است. در مجموعه داده MQTT براساس نتایج بست آمده در طبقه بندی چند کلاسه در کل بیشترین میزان پارامترها برای الگوریتم XG و RF می‌باشد. محققان باید بر معیارهای عملکرد مثبت و منفی کاذب تمرکز کنند که عملکرد مدل تشخیص نفوذ را کاهش می‌دهد، مطالعه تجربی نشان داده است که هیچ الگوریتم یادگیری ماشینی وجود ندارد که بتواند انواع حملات را به طور موثر تشخیص دهد. در آینده، ویژگی‌های مرتبط را می‌توان از مجموعه داده اصلی استخراج کرد تا زمان محاسبات را کاهش داده و میزان دقت را افزایش داد.

مراجع و منابع

[1] Zhang C, Jia D, Wang L, Wang W, Liu F, Yang A. Comparative research on network intrusion detection methods based on machine learning. Computers & Security. 2022 Jul 28:102861.
[2] Umer MA, Junejo KN, Jilani MT, Mathur AP. Machine learning for intrusion detection in industrial control systems: Applications, challenges, and recommendations. International Journal of Critical Infrastructure Protection. 2022 Feb 17:100516.
[3] Faghih Aliabadi H. A Hybrid Method for Intrusion Detection in the IOT. International Journal of Web Research. 2022 Dec 1;5(2):54-60.
[4] Singh G, Khare N. A survey of intrusion detection from the perspective of intrusion datasets and machine learning techniques. International Journal of Computers and Applications. 2022 Jul 3;44(7):659-69.
[5] Joshal KS, Gupta N, Tomar A. Internet of things-based smart grid: an overview. International Journal of Energy Technology and Policy. 2022;18(1):57-70.
[6] Jiang, J.R. Short Survey on Physical Layer Authentication by Machine-Learning for 5G-based Internet of Things. In Proceedings of the 2020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII), Kaohsiung, Taiwan, 21–23 August 2020; pp. 41–44
[7] S. Gurung, M. Kanti Ghose, and A. Subedi, “Deep Learning Approach on Network Intrusion Detection System using NSL-KDD Dataset, vol. 11, no. 3, pp. 8–14, 2019, doi: 10.5815/ijcnis.2019.03.02.
[8] Al-Marri NA, Ciftler BS, Abdallah MM. Federated mimic learning for privacy preserving intrusion detection. In2020 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom) 2020 May 26 (pp. 1-6). IEEE.
[9] Rey V, Sánchez PM, Celdrán AH, Bovet G. Federated learning for malware detection in iot devices. Computer Networks. 2022 Feb 26;204:108693.
[10] Y. Otoum, D. Liu, and A. Nayak, ‘‘DL-IDS: A deep learning-based intrusion detection framework for securing IoT,’’ Trans. Emerg. Telecommun. Technol., vol. 14, 2019
[11] https://research.unsw.edu.au/projects/unsw-nb15-dataset
[12] Hindy, H., Tachtatzis, C., Atkinson, R., Bayne, E., Bellekens, X.: MQTT-IOTIDS2020: MQTT internet of things intrusion detection dataset. IEEE Dataport (2020).