دومین کنفرانس ملی آخرین دستاوردهای مهندسی داده و دانش و محاسبات نرم

سیستم کشف نفوذ با استفاده از الگوریتم

تقویت گرادیان سبک

امیر حسین امیری[۱]، عبدالرضا رسولی کناری [۲]، مرتضی محجل کفشدوز[۳]

[۱] گروه کامپیوتر، دانشکده برق و کامپیوتر، دانشگاه صنعتی قم، قم
amiri.ah@qut.ac.ir
[۲] گروه کامپیوتر، دانشکده برق و کامپیوتر، دانشگاه صنعتی قم، قم
rasouli@qut.ac.ir
[۳] گروه کامپیوتر، دانشکده برق و کامپیوتر، دانشگاه صنعتی قم، قم
mohajjel@qut.ac.ir

چکیده

با توجه به افزایش روز افزون استفاده از شبکه‌های اینترنت اشیاء[۱] و کاربردهای گوناگون آن در زندگی روزمره خصوصا در بخش سلامت و ایمنی نفوذ و حملات به این شبکه‌ها نیز افزایش چشمگیری داشته است. این نفوذها و حملات در برخی از موارد می‌تواند خطرات و آثار جبران ناپذیری را ایجاد کند. در این مقاله ما سعی داریم تا سیستم کشف نفوذی[۲] مبتنی بر الگوریتم تقویت گرادیان سبک[۳] ارائه دهیم. سیستم مورد نظر ما به محدودیت‌های موجود در اجزاء این شبکه‌ها از نظر منابع محاسباتی و ذخیره‌سازی و لزوم شناسایی دقیق نفوذ، توجه دارد. ما این هدف را با بکارگیری الگوریتم‌های یادگیری ماشین[۴] که کاربردهای متنوعی دارند، دنبال می‌کنیم. در این خصوص تحقیقاتی صورت گرفته است و نتایج حاصله با دقت‌های مختلف بدست آمده است. در این تحقیق ما از مجموعه داده‌ای[۵] NSL-KDD و همچنین چند الگوریتم دیگر برای مقایسه عملکرد آنها با الگوریتم مورد نظر خودمان استفاده کردیم. نتایج حاصله از تحقیق نشان می‌دهد که الگوریتم تقویت گرادیان سبک در شناسایی حملات نسبت به الگوریتم‌های دیگر از عملکرد بهتری برخوردار می‌باشد.

کلمات کلیدی

سیستم کشف نفوذ، شبکه اینترنت اشیاء، الگوریتم یادگیری ماشین، الگوریتم تقویت گرادیان سبک

۱- مقدمه

امروزه اینترنت اشیاء که شامل مجموعه فراوانی از دستگاه‌ها، حسگرها، و ماشین‌آلاتی می‌باشد که از بستر اینترنت استفاده کرده و با یکدیگر ارتباط برقرار کرده و عملیات خاصی را انجام می‌دهند، بسیار مورد توجه قرار گرفته است. تعداد این دستگاه‌ها تا کنون بالغ بر ۲۵ بیلیون دستگاه می‌باشدکه از طریق اینترنت باسیم و بی‌سیم به یکدیگر متصل شده‌اند. تکنولوژی‌های بسیار متنوعی در شبکه اینترنت ا شیاء بکار گرفته شده ا ست و با توجه به پیشرفت روزافزون این تکنولوژی‌ها امکان اتصال هر دستگاه فیزیکی به اینترنت به زودی مهیا خواهد شد. این شبکه با توجه به حجم بالای د ستگاه‌های درگیر، داده‌های بسیار زیادی تولید می‌کند. این حجم زیاد از اطلاعات باعث بروز خطرات مختلفی می شود و این خطرات می‌تواند محرمانگی و امنیت اطلاعات کاربران را تهدید کند. اطمینان از امنیت داده‌های ذخیره شده و همچنین در حال انتقال داخل شبکه باید مورد توجه قرار گیرد. راه حل ایجاد امنیت باید علاوه بر امنیت دستگاه، امنیت کل سیستم را مورد توجه قرار دهد. از آنجایکه اعضای این شبکه می‌توانند متحرک با شند برای ارتباط با یکدیگر از اینترنت بی سیم ا ستفاده می‌کنند و این مو ضوع باعث می شود که حملاتی که در این نوع ارتباط وجود دارد نیز گریبان‌گیر این شبکه‌ها شود. برای اطمینان از امنیت اطلاعات در اکوسیستم شبکه اینترنت اشیاء راه حل‌هایی ارائه شده است ولی با توجه به محدودیت های سخت‌افزاری و محاسباتی موجود در اغلب دستگاه‌های اینترنت اشیاء چالش ایجاد می‌کند بنابراین برای ایجاد امنیت در

آنها باید از الگوریتم‌ها و روش‌هایی استفاده کرد که توان اجرایی شدن بر روی این سیستم‌ها را داشته باشند.

برای شناسایی نفوذ در شبکه اینترنت اشیاء از سیستم‌های کشف نفوذ (Intrusion Detection System) یا به صورت مخفف IDS استفاده می‌شود.

با توجه به ساختار لایه‌ای شبکه اینترنت اشیاء و لزوم ایجاد امنیت و جلوگیری از نفوذ مهاجمین به لایه‌های مختلف شبکه و جلوگیری از دسترسی آنها به داده‌های ارسالی از دستگاه‌ها و تجهیزات متصل به شبکه و ضعف این تجهیزات و دستگاه‌ها از نظر منابع محاسباتی، حافظه‌ای و ذخیره‌سازی نیاز به سیستم کشف نفوذی می‌باشد که دارای پیچیدگی زمانی و فضایی کمتری بوده و با استفاده از منابع کم این دستگاه‌ها توان کشف نفوذ را داشته باشد. همچنین سرعت و دقت در شناسایی حملات و اعلان آن برای جلوگیری از خطرات ناشی از نفوذ نیز دارای اهمیت خاصی می‌باشد که بایستی در سیستم مورد نظر دیده شود. اگر مهاجمی بتواند اختیار بخش یا بخش‌هایی از شبکه را بعهده بگیرد، خصوصا در شبکه‌های سلامت و پزشکی، می‌تواند خطرات جبران ناپذیری ایجاد کند. لذا می‌توان چالش‌های اصلی در این زمینه را بصورت موارد زیر مطرح کرد:

۱) افزایش کارایی

۲) رعایت کمبود منابع در اجزاء شبکه اینترنت اشیاء

در این تحقیق تلاش داریم تا با استفاده از الگوریتم LightGBM که از الگوریتم‌های ترکیبی از دسته الگوریتم‌های ناظر[۶] می‌باشد ولی با توجه به ساختارش، سبک بوده، و می‌تواند بر کمبود منابع در اجزاء شبکه اینترنت اشیاء غلبه کرده و نفوذها به اینگونه شبکه‌ها را شناسایی کند، سیستم کشف نفوذ ارائه دهیم. همچنین سعی می‌کنیم با شناسایی ویژگی‌هایی[۷] که تاثیر بیشتری بر پیش‌بینی نفوذ دارند، دقت شناسایی را بالا ببریم و با کاهش ویژگی‌های اضافی از پیچیدگی الگوریتم بکاهیم.

بطور کلی می‌توان اهداف این تحقیق را بصورت زیر بیان کرد:

۱) بالا بردن دقت شناسایی نفوذ با کاهش ویژگی‌ها و بکارگیری ویژگی‌هایی که دقت تشخیص نفوذ را بالا می‌برد

۲) رعایت کمبود منابع در اجزای شبکه اینترنت اشیاء با استفاده از LightGBM که منابع کمتری می‌خواهد

در ادامه مقاله در بخش دوم به مقالاتی که در این زمینه منتشر شده است می‌پردازیم و به فرآیند صورت گرفته در آنها اشاره می‌کنیم. بخش سوم با پیش‌زمینه‌های مورد نیاز از جمله ساختار و تنوع سیستم‌های کشف نفوذ، الگوریتم‌های یادگیری ماشین و خصوصا الگوریتم تقویت گرادیان سبک بیشتر آشنا می‌شویم. در بخش چهارم روش پیشنهادی تحقیق را توضیح می‌دهیم و در رابطه با داده‌های مورد آزمایش، ابزارهای مورد استفاده، متغیرها و روش‌های تحلیل نتایج تحقیق، بیشتر توضیح می‌دهیم. در بخش پنجم نتیجه حاصل از تحقیق را مورد مطالعه قرار می‌دهیم. ارزیابی و تحلیل نتایج حاصله از تحقیق و مقایسه نتایج بدست آمده با نتایج تحقیقات قبلی را در این بخش خواهیم داشت و نهایتا در بخش ششم خلاصه‌ای از تحقیق ارائه خواهیم کرد.

۲- کارهای مرتبط

مطالعه بر روی امنیت در اینترنت اشیاء روی مدل‌های حمله بصورت جداگانه و با استفاده از تکنیک‌های یادگیری ماشین تمرکز دارد. برای مثال تشخیص نفوذ در تعداد زیادی از مطالعات مورد توجه قرار گرفته است. روش کار در اکثر مطالعات انجام شده قبلی بصورت تجربی می‌باشد.

در مرجع [۱] در سال ۲۰۲۰ سیستم کشف نفوذی معرفی شده که از LightGBM جهت انتخاب ویژگی‌ها و از Autoencoder برای طبقه‌بندی[۸] استفاده کرده است. او از توانایی الگوریتم LightGBM استفاده کرده و ویژگی‌هایی که در فرآیند طبقه‌بندی موثرترند را انتخاب کرده است. سپس با استفاده از یک Autoencoder بخش طبقه‌بندی را انجام داده است. او روش خود را در محیط پایتون و بر روی بانک اطلاعات NSL-KDD اجرا کرده و به دقت ۸۹,۸۲ درصد دست یافت.

در مرجع [۲] در سال ۲۰۲۰ مدلی برای شناسایی حملات در شبکه معرفی شده که برپایه الگوریتم XGBoost و بکارگیری GPU[۹] بوده است. او و همکارانش از یک الگوریتم ترکیبی تجزیه و تحلیل مولفه‌های اصلی (PCA)[۵] بنام کرم شب تاب (Firefly) برای کشف حملات استفاده کرده‌اند. او برای کاهش ابعاد در این تحقیق از الگوریتم کرم شب تاب استفاده کرد و شناسایی حملات را با الگوریتم XGBoost انجام داد.

در مرجع [۳] در سال ۲۰۱۷ با استفاده از ترکیب الگوریتم‌های کلونی زنبور عسل و AdaBoost سیستم کشف نفوذ معرفی کرد. او و همکارانش در سیستمی که ارائه دادند از الگوریتم کلونی زنبور عسل بعنوان انتخاب کننده ویژگی‌ها و از الگوریتم AdaBoost برای طبقه‌بندی استفاده کردند. او سیستم خود را بر روی بانک‌های اطلاعاتی NSL-KDD و ISCXIDS2012 آزمایش کرد. برای انجام سناریو اول از بانک اطلاعات NSL-KDD استفاده کرد و برای هر حمله دقت شناسایی متفاوتی بدست آورد. بالاترین دقت شناسایی برای حمله DoS[۱۰] بوده که معادل ۹۹,۸۶ درصد بود. برای انجام سناریوی دوم از بانک اطلاعاتی ISCXIDS2012 استفاده کرد و دقت بدست آمده معادل ۸۳ درصد بود.

در مرجع [۴] در سال ۲۰۲۰ بر روی سه نوع از الگوریتم‌های تقویتی بنام‌های Real AdaBoost، Gentel AdaBooost و Modest AdaBoost کارایی، دقت و سرعت مقایسه شد. او برای انجام این تحقیق از ۵ بانک اطلاعاتی مختلف استفاده کرد و هر سه الگوریتم را بر روی آنها آزمایش کرد. بانک‌های اطلاعاتی بکار گرفته شده عبارتند از KDD Cup99، UNSW-NB15، TRAbID، NSL-KDD و CICIDS2017. او و همکارانش با توجه به نتایج بدست آمده متوجه شدند که کارایی Gentel و Real AdaBoost مشابه هم بوده و Modest AdaBoost کارایی کمتری داشته است ولی Modest AdaBoost، ۷ درصد سریعتر عمل کرده است.

در مرجع [۵] در سال ۲۰۱۹ مقایسه‌ای بر کارایی الگوریتم Adaptive Boost در سیستم کشف نفوذ در شبکه انجام شد. او و همکارانش ترکیبی از الگوریتم Boosting به همراه سه الگوریتم طبقه بندی ضعیف (بیز ساده[۱۱]، درخت تصادفی[۱۲] و بیز ساده مبتنی بر ویژگی‌های وزن دار مرتبط[۱۳]) را بکار گرفت و آنها را بر روی دو بانک اطلاعاتی NSL-KDD و KAGGLE آزمایش کرد. او و همکارانش با بکار گیری تکنیک Boosting بر روی الگوریتم‌های ضعیف توانستند نتایج خوبی بدست آورند. بهترین نتیجه در تحقیق او مربوط به ترکیب Adaptive Boosting و Random Tree می‌باشد که معادل ۹۸,۴۵ درصد دقت در تشخیص ناهنجاری در ترافیک شبکه بود.

در مرجع [۶] در سال ۲۰۱۹ سه الگوریتم C5.0، AdaBoost.M1 و AdaBoost.SAMME را بر روی بانک اطلاعاتی کارایی دانش آموزان پیاده

سازی کرد. کارایی دانش آموزان یک عامل مهم در رقابت موسسات آموزشی می‌باشد. برای توسعه این رقابت نیاز به پیش‌بینی کارایی دانش آموزان می‌باشد. در سناریوی اول برای مقایسه عملکرد الگوریتم‌ها از اعتبار سنجی متقابل ۱۰ قسمتی[۵]استفاده کرد. خروجی این سناریو نشان داد که AdaBoost.M1 و AdaBoost.SAMME در طبقه‌بندی دودویی مانند روش پایه عمل می‌کنند. سناریوی دوم برای ارزیابی الگوریتم‌های تقویتی بر اساس تعداد متفاوت داده‌های آموزشی طراحی شد. در این سناریو AdaBoost.M1 کارایی بهتری داشت. در سناریوی سوم برای آموزش از یک بانک اطلاعاتی استفاده شد ولی برای پیش‌بینی داده‌های بانک دیگری بکار گرفته شد.

در مرجع [۷] در سال ۲۰۲۰ از الگوریتم XGBoost در سیستم کشف حمله ترکیبی استفاده شد. او در تحقیق خود از بانک اطلاعاتی KDD-Cup99 استفاده کرد و دقت حاصله از بکارگیری روش او ۹۹,۹۵ درصد بوده است. او و همکارانش با استفاده از این روش امکان شناسایی حملات ناشناخته توسط سیستم کشف نفوذ خود را ایجاد کردند. استفاده از روش‌های ترکیبی این امکان را به این تحقیق داده است که کارآمدتر بوده و تنها به شناسایی حملات خاصی منحصر نشود.

۳- پیش زمینه

دراین بخش به تعاریف اولیه می‌پردازیم و اطلاعات مورد نیاز در رابطه با مقاله را ارائه می‌دهیم.

۳-۱- شبکه اینترنت اشیاء و تهدیدات آن

شبکه اینترنت اشیاء شامل میلیون‌ها شیء که توسط بستر اینترنت (با سیم و بی‌سیم) با هم در ارتباط بوده و برای هم داده ارسال می‌کنند، می‌باشد. این داده‌ها مهمترین محل ورود مهاجمین به اینگونه شبکه‌ها می‌باشند. با توجه به ساختار لایه‌ای شبکه اینترنت اشیاء حملات و خطراتی که در هر لایه وجود دارد با لایه‌های دیگر متفاوتند و برای شناسایی و کشف آنها باید از راه‌کارهای مختلفی استفاده شود. یکی از روش‌های دسته‌بندی لایه‌های شبکه اینترنت اشیاء بدین ترتیب می‌باشد: لایه حسگرها[۱۶]لایه شبکه[۱۷]لایه میانی[۱۸] Gateway و لایه کاربرد[۱۹]شکل (۱) نمایش تعدادی از حملات موجود در لایه‌های مختلف این شبکه می‌باشد.

لایه حسگرها در واقع لایه فیزیکی شبکه اینترنت اشیاء می‌باشد که شامل حسگرها و عمل‌کننده‌ها می‌باشد. انواع مختلف حسگر برای دریافت انواع داده‌ها وجود دارد بعنوان مثال می‌توان به حسگر فراصوت، حسگر دوربین، حسگر تشخیص دود، حسگر دما و رطوبت اشاره کرد. فناوری‌های مختلفی در این لایه بکار می رود مانند RFID، GPS، RSN، WSN و غیره. برخی از انواع تهدیدهایی که در این لایه وجود دارد عبارتند از :

- Node Capturing
- Malicious Code Injection Attack
- False Data Injection Attack
- Side Channel Attacks
- Eavesdropping and Interference
- Sleep Deprivation Attack
- Booting Attack

لایه شبکه کار انتقال اطلاعات ایجاد شده توسط حسگرها به بخش‌های محاسباتی را انجام می‌دهد. موضوعات امنیتی این لایه عبارتند از :

- Phishing Site Attack
- Access Attack
- DoS/DDoS Attack
- Data Transit Attack
- Routing Attack

هدف اصلی لایه میانی در شبکه اینترنت اشیاء ایجاد یک لایه جداکننده بین لایه شبکه و لایه نرم‌افزار می‌باشد. لایه میانی همچنین توانایی مهیا نمودن منابع ذخیره سازی و محاسباتی را نیز دارد. لایه میانی شامل واسطها، محل ذخیره پایدار داده‌ها، سیستم‌های صف، یادگیری ماشین و موارد دیگر می‌باشد. برخی از حملات احتمالی در لایه میانی عبارتند از :

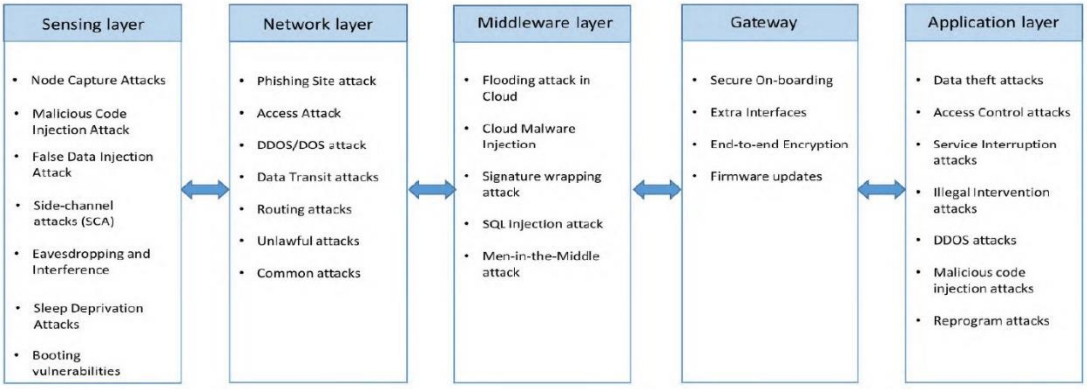
- Man in the Middle Attack
- SQL Injection Attack
- Signature Wrapping Attack
- Cloud Malware Injection
- Flooding Attack in Cloud

دروازه[۲۰]لایه وسیعی است که وظیفه مهمی در اتصال چندین دستگاه، فرد، شیء و خدمات ابری را بعهده دارد. دروازه‌ها برای رمزگشایی و رمزنگاری داده‌های شبکه اینترنت اشیاء و ترجمه پروتکل‌ها برای ارتباط بین لایه‌ها استفاده می‌شوند. امروزه سیستم شبکه اینترنت اشیاء ناهمگون می‌باشد که شامل LoraWan، ZigBee، Z-Wave و پشته های TCP/IP با دروازه‌های فراوانی بین آنها می‌باشد. برخی از چالش‌های امنیتی دروازه‌ها عبارتند از:

- Source On-boarding
- Extra Interfaces
- End-to-End Encryption
- Firmware Update

لایه نرم‌افزار (کاربرد) مستقیما با کاربر نهایی در ارتباط می‌باشد و خدمت را برای او مهیا می‌کند. نرم‌افزارهای شبکه اینترنت اشیاء مانند خانه هوشمند، اندازه‌گیری هوشمند، شبکه هوشمند، شهر هوشمند و غیره در این لایه قرار دارند. موضوعات امنیتی اصلی در این لایه عبارتند از:

- Data Thefts
- Access Control Attack
- Service Interruption Attack
- Malicious Code Injection Attack
- Sniffing Attack
- Reprogram Attack



شکل (۱) : حملات و خطرات شبکه اینترنت اشیاء به تفکیک لایه‌ها

در این تحقیق ما سعی داریم در رابطه با چالش‌های موجود در لایه شبکه سیستم کشف نفوذ ارائه دهیم.

۲-۳- یادگیری ماشین و امنیت شبکه اینترنت اشیاء

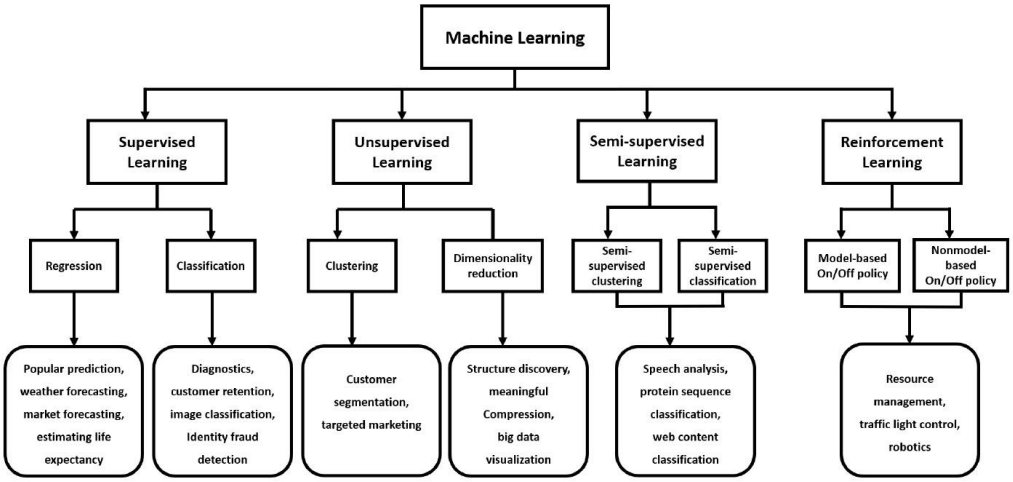
یادگیری ماشین (Machine Learning) که بصورت مخفف ML گفته می‌شود، مطالعه درباره الگوریتم‌های رایانه‌ای می‌باشد که خودکار شدن فرآیندها، یادگیری و پیش بینی را فراهم می‌کند. یادگیری ماشین یکی از شیوه‌های هوش مصنوعی می‌باشد که ماشین‌ها را با الگوریتم‌های مختلفی آموزش می‌دهد و به دستگاه‌ها کمک می‌کند تا بجای برنامه‌ریزی از قبل، از تجربیاتشان برای یادگیری استفاده کنند. [۸] بیان می‌کند یادگیری ماشین نیازی به کمک انسان، در حل معادلات پیچیده ریاضی ندارد، و می‌تواند در شبکه‌های پویا عمل کند. [۹] می‌گوید در هر حال روش‌های یادگیری ماشین می‌تواند برای کشف زود هنگام حملات مختلف در شبکه اینترنت اشیاء با بررسی رفتار دستگاه‌ها بکار گرفته شود. علاوه بر این، برای مقابله با کمبود منابع در دستگاه‌های شبکه اینترنت اشیاء، می‌توان با انتخاب الگوریتم‌های یادگیری ماشین مختلف راه حل مناسب را پیدا کرد. یکی از کاربردهای الگوریتم‌های یادگیری ماشین در سیستم‌های کشف نفوذ می‌باشد. شکل (۲) دسته بندی کلی الگوریتم‌های یادگیری ماشین‌و برخی کاربردهای آنها را نمایش می‌دهد. دسته‌بندی کلی برای الگوریتم‌های یادگیری ماشین عبارتند از یادگیری باناظر، یادگیری بدون‌ناظر، یادگیری نیمه‌ناظر و یادگیری تقویت شده. از آنجاییکه در شناسایی حملات و نفوذهای صورت گرفته در شبکه اینترنت اشیاء اطلاعات و مشخصات انواع حملات صورت گرفته در دست می‌باشد و می‌توان آنها را به عنوان شاخص مورد استفاده قرارداد، می‌توان از الگوریتم‌هایی که در دسته یادگیری باناظر می‌باشند استفاده کرد.

همچنین چون هدف از شناسایی در واقع دسته‌بندی و تعیین نوع حمله می‌باشد باید از الگوریتم‌های دسته‌بندی (Classifier) استفاده کرد. برخی از این الگوریتم‌ها در شکل (۳) نشان داده شده‌اند.

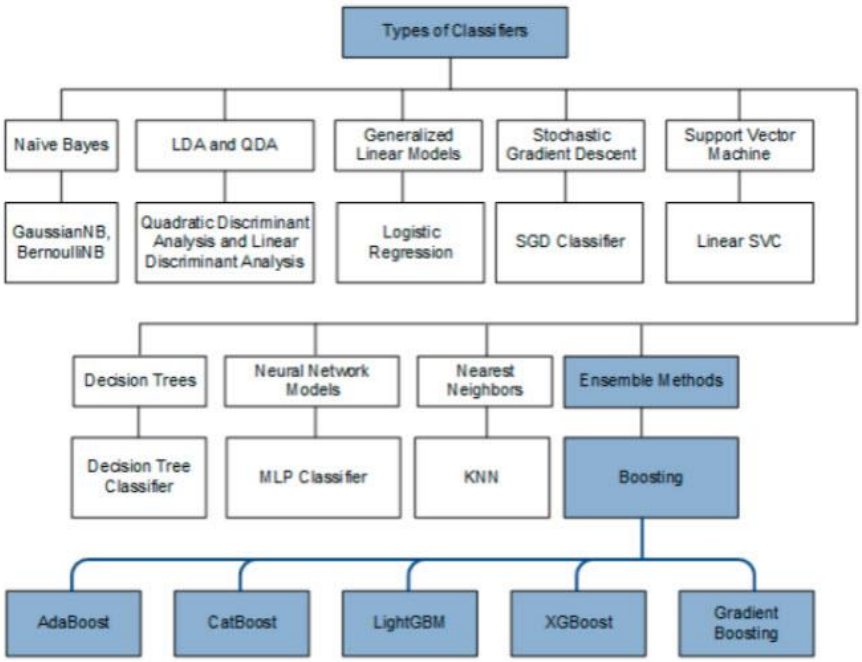
الگوریتم تقویت گرادیان سبک (LightGBM) یک چهارچوب تقویت گرادیان سریع، توزیع شده، با کارایی بالا بر اساس الگوریتم‌های درخت تصمیم می‌باشد که برای رتبه‌بندی، دسته‌بندی و بسیاری وظایف دیگر یادگیری ماشین استفاده می‌شود. LightGBM یک چهارچوب تقویت گرادیان می‌باشد که سه الگوریتم یادگیری پایه را استفاده می‌کند. این چهارچوب برای اینکه بصورت موثر و توزیع شده باشد با مزایای زیر طراحی شده است:

- سرعت آموزش و بهره‌وری بیشتر
- نیاز به حافظه کمتر
- دقت بالاتر
- پشتیبانی از موازی‌کاری و یادگیری از طریق GPU
- توانایی پشتیبانی از داده‌های با حجم بالا

چند سال پیش، شرکت میکرو سافت چهارچوب تقویت گرادیان خود، LightGBM را معرفی کرد. LightGBM ۶ بار سریعتر از XGBoost می‌باشد. LightGBM نسبتا یک الگوریتم جدید می‌باشد و یک لیست طولانی از متغییرها دارد که در سند LightGBM آورده شده است.



شکل (۲) : دسته بندی کلی الگوریتم‌های یادگیری ماشین و برخی کاربردهای آنها

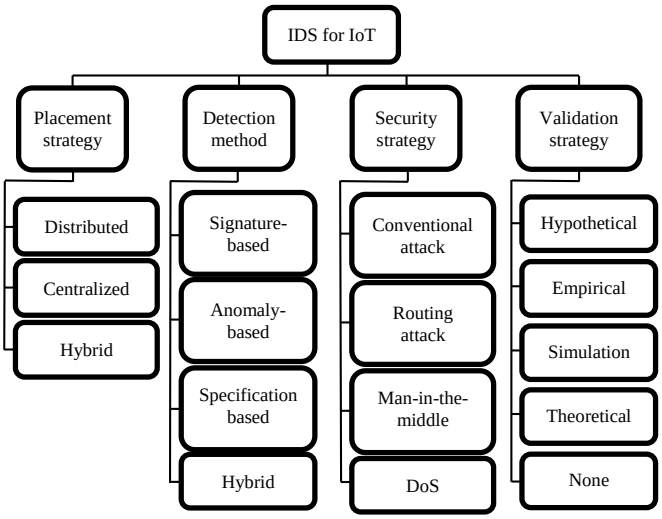


شکل (۳) : انواع الگوریتم های دسته بندی

حجم مجموعه‌های داده‌ای بسرعت افزایش می‌یابند. برای الگوریتم‌های علوم داده سنتی ایجاد نتایج دقیق بر روی این مجموعه‌های داده‌ای بسیار مشکل می‌باشد. LightGBM می‌تواند داده‌های با حجم زیاد را با استفاده از حافظه کمتر پشتیبانی کند. LightGBM همچنین از یادگیری توسط GPU پشتیبانی می‌کند و به این علت دانشمندان داده بطورگسترده‌ای از آن در توسعه ابزارهای علوم داده استفاده می‌کنند.

۳-۳- سیستم‌های کشف نفوذ در شبکه اینترنت اشیاء

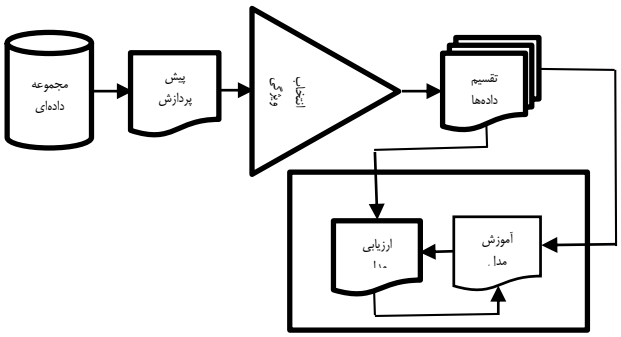
جلوگیری از نفوذ و شناسایی به موقع آن لازمه اصلی ایجاد امنیت در این شبکه‌ها می‌باشد. این وظیفه را سیستم‌های کشف نفوذ (IDS) به عهده دارند. سیستم‌های کشف نفوذ به روش‌های مختلفی دسته‌بندی می‌شوند. یکی از روش‌های دسته‌بندی سیستم‌های کشف نفوذ با توجه به محل قرارگیری در اکوسیستم شبکه اینترنت اشیاء، روش شناسایی حملات، نوع خطر قابل شناسایی و سیاست اعتبار سنجی می‌باشد که در شکل (۴) نشان داده شده است.



شکل (٤) : دسته‌بندی سیستم‌های کشف نفوذ

٤- روش پیشنهادی

معماری طرح پیشنهادی در شکل (۵) آورده شده است.

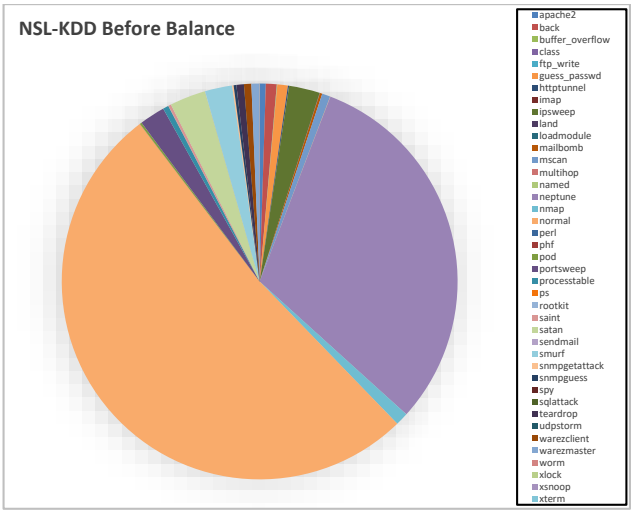


شکل (۵) : معماری طرح پیشنهادی

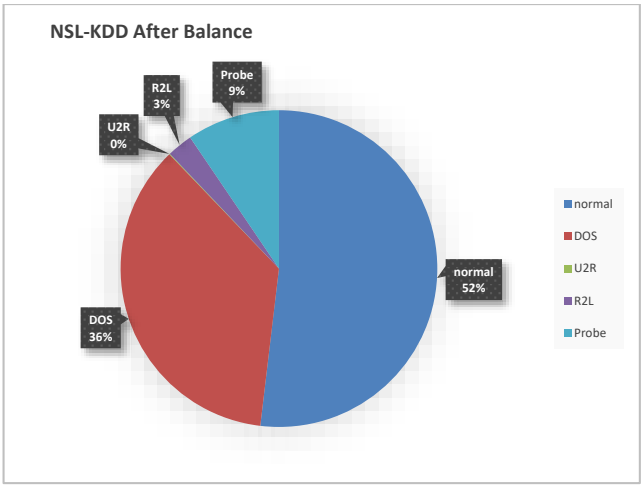
برای پیاده سازی الگوریتم مورد نظر خودمان از مجموعه داده‌ای NSL-KDD اســتفاده کردیم و الگوریتم تقویت گرادیان ســبک را با هر کدام از الگوریتم‌هایی که قبلا بر روی این مجموعه داده‌ای تســت شــده بود مقایســه کردیم.

٤-۱- پیش پردازش و انتخاب ویژگی

قبل از اســتفاده از مجموعه داده‌ای بخاطر عدم توازن در داده‌های آن، ابتدا داده‌ها را به دسته‌های کلی حملات دسته‌بندی کردیم. در شکل (۶) پراکندگی داده‌ها قبل ازد سته‌بندی نشان داده شده است. همانطور که مشخص است درصــد برخی از حملات بســیار ناچیز می‌باشــد که در روند آموزش و یادگیری مدل اخلال ایجاد خواهد کرد بنابراین با دسته‌بندی داده‌ها این پراکندگی را در شکل (۷) به حداقل رساندیم. ما انواع حملات موجود در مجموعه داده‌ای اولیه که شــامل ۴۱ نوع می‌باشــد را به ۵ دســته کلی‌تر که شــامل داده‌های طبیعی، DoS، U2R[۲۲]، R2L[۲۱] و Probe تبدیل کردیم.



شکل (٦) : پراکندگی داده‌ها قبل از دسته بندی



شکل (٧) : پراکندگی داده‌ها بعد از دسته‌بندی

سپس با استفاده از ابزار StandardScaler داده‌ها را مقیاس‌بندی کرده و با استفاده از ابزار SelectKBest پایتون تعداد ویژگی‌ها را کاهش دادیم و بهترین ویژگی هایی که برای انجام عملیات آموزش مدل مناسب می‌باشند را انتخاب کردیم تا بتوانیم در مصرف حافظه و قدرت محاسباتی صرفه جویی کنیم.

٤-٢- مجموعه داده‌ای

مجموعه داده‌ای که در این مقاله مورد آزمایش قرار گرفت مجموعه داده‌ای NSL-KDD می‌باشد. در این مجموعه داده‌ای مشکلات موجود در مجموعه داده‌ای KDD-Cup99 برطرف گردیده است. این نسخه همچنان یک شبکه واقعی را نمایش نمی‌دهد، ولی بخاطر داشتن دریایی از داده‌ها برای سیستم‌های کشف نفوذ بر پایه شبکه می‌تواند بعنوان مجموعه داده‌ای مرجع مفید باشد و برای مقایسه روش‌های کشف نفوذ توسط محققین استفاده شود.

بعلاوه، تعداد داده‌های مجموعه آموزش و آزمون NSL-KDD مناسب می‌باشد. این مزیت این مجموعه داده‌ای را بدون نیاز به انتخاب تصادفی بخش کوچکی برای انجام آزمایشات مناسب می‌سازد. بنابراین، نتایج ارزیابی کارهای تحقیقاتی متفاوت یکدست و قابل مقایسه خواهند بود.

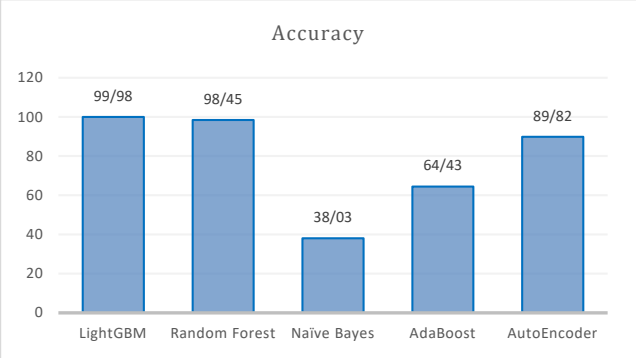
٤-٣- ابزارهای مورد استفاده

برای پیاده سازی آزمایشات خود از محیط نرم‌افزار پایتون و بسته‌های موجود در آن استفاده کردیم. همچنین از نرم‌افزار میکروسافت اکسل برای تهیه نمودارهای نتایج حاصله استفاده کردیم.

٥- ارزیابی نتایج

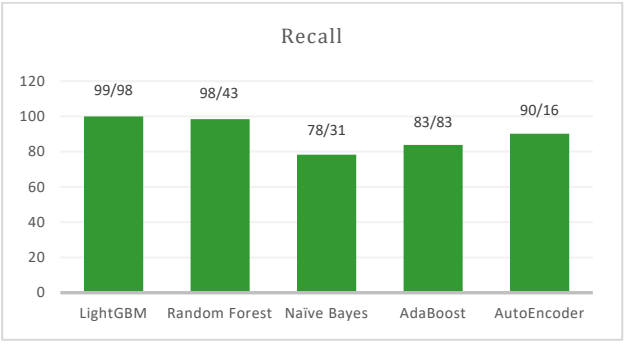
ما بعد از انجام عملیات پیش‌پردازش و انتخاب ویژگی‌های مناسب برای انجام عملیات آموزش و سپس پیش‌بینی بر روی مجموعه داده‌ای مورد استفاده در این تحقیق نتایج حاصله را در نمودارهایی به نمایش درآوردیم. برای مقایسه عملکرد الگوریتم‌های متفاوت با الگوریتم LightGBM از معیارهای Accuracy، Precision، Recall و F1-Score استفاده کردیم.

در شکل (٨) مقایسه میزان دقت بدست آمده توسط الگوریتم‌های بکار رفته در تحقیقات دیگر با الگوریتم مورد نظر ما بر روی مجموعه داده‌ای NSL-KDD در پیش‌بینی نشان داده شده است. می‌توان مشاهده کرد که الگوریتم مورد نظر ما نسبت به دیگر الگوریتم‌های مورد آزمون عملکرد مناسبتری داشته و دقت بالاتری را ثبت کرده است.



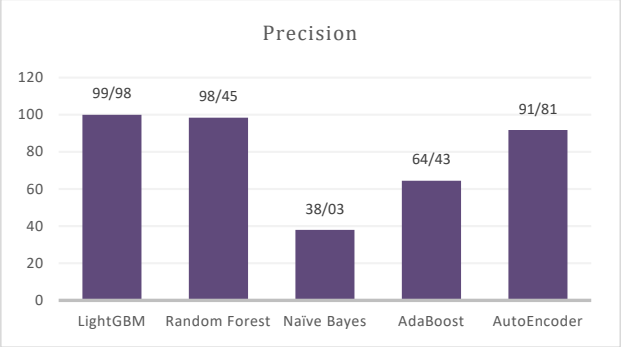
شکل (٨) : نمودار مقایسه دقت بر روی مجموعه داده‌ای NSL-KDD

در شکل (٩) مقایسه معیار Recall بین الگوریتم LightGBM و دیگر الگوریتم‌ها بر روی مجموعه داده‌ای NSL-KDD نمایش داده شده است.

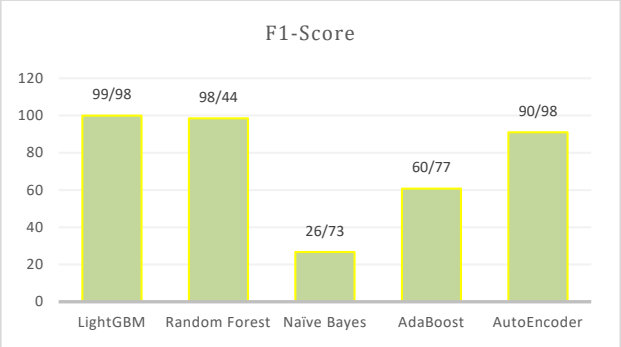


شکل (٩) : نمودار مقایسه Recall بر روی مجموعه داده‌ای NSL-KDD

در شکل‌های (١٠،١١) مقادیر مربوط به Precision و F1-Score آورده شده است. همانطور که در تمامی نمودارها مشاهده می‌شود الگوریتم انتخابی ما بالاترین درصد را دارا می‌باشد.



شکل (۱۰) : نمودار مقایسه Precision بر روی مجموعه داده‌ای NSL-KDD



شکل (۱۱) : نمودار مقایسه F1-Score بر روی مجموعه داده‌ای NSL-KDD

٦- نتیجه

در این مقاله، ما از الگوریتم تقویت گرادیان سبک برای بالا بردن کارایی کشف حملات در شبکه اینترنت اشیاء استفاده کردیم و مشخص شد که بر روی مجموعه داده‌ای NSL-KDD پاسخ خوبی می‌دهد. در کارهای بعدی سعی خواهیم کرد این الگوریتم را در مجموعه‌های داده‌ای دیگر خصوصا مجموعه‌های داده‌ای که بصورت واقعی ایجاد شده باشند بکار بگیریم و همچنین بتوانیم راه حلی برای شناسایی دقیق نوع حمله بصورت جزئی‌تر بیابیم.

مراجع

[1] C. Tang, N. Luktarhan, and Y. Zhao, "An Efficient Intrusion Detection Method Based on LightGBM and Autoencoder," *Symmetry*, vol. 12, no. 9, 2020.

[2] S. Bhattacharya *et al.*, "A Novel PCA-Firefly Based XGBoost Classification Model for Intrusion Detection in Networks Using GPU," *Electronics*, vol. 9, no. 2, 2020.

[3] M. Mazini, B. Shirazi, and I. Mahdavi, "Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms," *Journal of King Saud University - Computer and Information Sciences*, vol. 31, no. 4, pp. 541-553, 2019.

[4] A. Shahraki, M. Abbasi, and Ø. Haugen, "Boosting algorithms for network intrusion detection: A comparative evaluation of Real AdaBoost, Gentle AdaBoost and Modest AdaBoost," *Engineering Applications of Artificial Intelligence*, vol. 94, 2020.

[5] S. Sivanantham, R. Abirami, and R. Gowsalya, "<Comparing the Performance of Adaptive Boosted Classifiers in Anomaly based Intrusion Detection System for Networks.pdf>," 2019.

[6] F. Jauhari and A. A. Supianto, "Building student's performance decision tree classifier using boosting algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, 2019.

[7] B. S. Bhati, G. Chugh, F. Al - Turjman, and N. S. Bhati, "An improved ensemble based intrusion detection technique using XGBoost," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 6, 2020.

[8] M. I. Jordan and T. J. S. Mitchell, "Machine learning: Trends, perspectives, and prospects," vol. 349, pp. 255 - ۲۶۰, ۲۰۱۵.

[9] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "<Machine Learning in Wireless Sensor Networks-Algorithms, Strategies, and Applications.pdf>," 2015.

زیرنویس‌ها

[1] Internet of Things (IoT)
[2] Intrusion Detection System (IDS)
[3] Light Gradient Boosting Machine (LightGBM)
[4] Machin Learning (ML)
[5] Data Set
[6] Supervised Learning
[7] Feature
[8] Classifier
[9] Graphic Processing Unit
[10] Principal Component Analysis
[11] Denial of Service
[12] Naïve Bayes
[13] Random Tree
[14] Weighted Naïve Bayes
[15] 10-fold cross-validation
[16] Sensing Layer
[17] Network Layer
[18] Middleware Layer
[19] Application Layer
[20] Gateway
[21] User to Root Attacks
[22] Remote to User Attacks