سامانه‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)

Summer 2024, 5 (2), 71-80
DOR:

Received: 16 Apr 2024
Accepted: 4 Jun 2024

مقاله پژوهشی

Hybrid Intrusion Detection System: Leveraging XGBoost, Genetic Algorithms and k-Means Clustering

Ladan Riazi*

Assistant Professor, Department of Information Technology Management, Science and Research Branch, Islamic Azad University, Tehran, Iran. **Corresponding Author*, ladan.riazi@srbiau.ac.ir

Abstract

Introduction: The proliferation of internet applications has led to an interconnected digital landscape where network security is a paramount concern. Intrusion detection systems (IDS) play a critical role in safeguarding networks by identifying and mitigating unauthorized access and malicious activities. This research proposes a novel IDS framework that integrates XGBoost, k-Means clustering, and genetic algorithms to enhance intrusion detection capabilities.

Method: The proposed IDS framework commences with data preprocessing to ensure compatibility with machine learning algorithms. Subsequently, k-Means clustering is employed to group unlabeled data, generating valuable insights that augment the dataset. To optimize feature selection, a genetic algorithm is integrated into the framework. This algorithm iteratively evaluates feature subsets using XGBoost to identify the most informative features for intrusion detection.

Results: The efficacy of the proposed IDS was evaluated using the NSL-KDD dataset. Comparative analysis with established methods, including decision trees, XGBoost, KNN, and random forests, demonstrated the superior performance of the proposed approach in terms of precision, recall, and F1-score.

Discussion: The integration of XGBoost, k-Means clustering, and genetic algorithms in the proposed IDS framework offers significant advantages. k-Means clustering provides valuable insights into unlabeled data, while genetic algorithms enable efficient feature selection. The empirical results underscore the effectiveness of the proposed approach in accurately detecting intrusions in computer networks.

Keywords: Intrusion Detection, Computer Networks, XGBoost, Clustering, k-Means, Genetic Algorithm.

سامانه‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)
انجمن علمی تجارت الکترونیکی ایران
واحد زنجان

سیستم تشخیص نفوذ ترکیبی با استفاده از الگوریتم‌های XGBoost، ژنتیک و خوشه‌بندی k-Means

تاریخ دریافت: ۲۸/ ۰۱/ ۱۴۰۳
تاریخ پذیرش: ۱۵/ ۰۳/ ۱۴۰۳

دورهٔ پنجم، تابستان ۱۴۰۳
شمارهٔ دوم، صص: ۷۱-۸۰

لادن ریاضی*

استادیار، گروه مدیریت فناوری اطلاعات، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران. (نویسندهٔ مسئول)
ladan.riazi@srbiau.ac.ir

چکیده: گسترش برنامه‌های کاربردی اینترنتی منجر به ایجاد یک فضای دیجیتالی به هم پیوسته شده‌است که در آن امنیت شبکه یک نگرانی اساسی است. سیستم‌های تشخیص نفوذ (IDS) با شناسایی و کاهش دسترسی‌های غیرمجاز و فعالیت‌های مخرب، نقش مهمی در حفاظت از شبکه‌ها ایفامی‌کنند. در این تحقیق یک چارچوب جدید برای سیستم‌های مذکور با استفاده از ادغام الگوریتم‌های XGBoost، خوشه‌بندی k-Means و الگوریتم‌های ژنتیک پیشنهاد شده‌است. چارچوب IDS پیشنهادی با پیش‌پردازش داده‌ها برای اطمینان از سازگاری با الگوریتم‌های یادگیری ماشین آغازمی‌شود. سپس، خوشه‌بندی k-Means برای گروه‌بندی داده‌های بدون برچسب اعمال-می‌شود و درک مناسبی از مجموعه داده حاصل‌می‌گردد. برای بهینه‌سازی انتخاب ویژگی، یک الگوریتم ژنتیک نیز در چارچوب پیشنهادی گنجانده‌شده‌است. این الگوریتم به‌طور مکرر زیر مجموعه‌های ویژگی را با استفاده از الگوریتم XGBoost ارزیابی‌می‌کند تا ارزشمندترین ویژگی‌ها را برای تشخیص نفوذ شناسایی‌کند. کارایی IDS پیشنهادی با استفاده از مجموعه داده NSL-KDD ارزیابی‌شد. در تجزیه و تحلیل مقایسه‌ای صورت‌گرفته با سایر روش‌های موجود، از جمله درختان تصمیم، XGBoost، KNN و جنگل‌های تصادفی، رویکرد پیشنهادی عملکرد برتری را از نظر معیارهای Precision ، Recall و F1-Score نشان‌داد.

واژه‌های کلیدی: تشخیص نفوذ در شبکه‌های کامپیوتری، الگوریتم XGBoost، خوشه‌بندی، الگوریتم K-Means.

۱. مقدمه

روند روبه‌رشد استفاده از شـــبکه‌های کامپیوتری به‌خصـــوص اینترنت، مهارت روبه‌رشد کاربران و مهاجمان این شـــبکـه‌هـا و وجود نقاط آسیب‌پذیری مختلف در نرم‌افزارها، موجب‌شـده‌اسـت که ایمن‌سـازی سیستم‌ها و شبکه‌های کامپیوتری، نسبت به گذشته از اهمیت بیشتری برخوردار باشـــد. وقتی یک سیستم کامپیوتری به یک شبکه متصل-می‌شود، در معرض ریسک بالایی قرارمی‌گیرد. ازجمله تهدیداتی که برای یک سیستم کامپیوتری وجوددارد، ویروس‌ها و نفوذها هستند [۱].

ظهور دستگاه‌های جدیدی که از اینترنت به‌عنوان ابزار اصلی خود استفاده‌می‌کنند، امنیت شبکه را به یکی از مهم‌ترین حوزه‌های امروزی در جهان تبدیل‌کرده‌است. نفوذ در محیط محاسباتی بسیار رایج است و باعث ایجاد مشکلات زیادی برای افراد، مشاغل خصوصی یا دولتی می‌شود که فکرمی‌کنند اطلاعات خود را خصوصی نگه‌می‌دارند. درنتیجه، امنیت داده‌ها به یک موضوع برجسته تبدیل‌شده‌است که باید به شیوه‌ای فوری موردتوجه قرارگیرد [۲]. روزانه میلیون‌ها رایانه قربانی رویه‌ای می‌شوند که باعث‌می‌شود کسب‌وکارها با افشای اطلاعات خصوصی‌شان در معرض رقابت‌های شرکتی، سرمایه و اعتبار خود را از دست بدهند. درنتیجه، امنیت شبکه برای کاربران کامپیوتر و سازمان‌ها بسیار مهم شده‌است. برای جلوگیری از دسترسی غیرمجاز به سیستم‌ها، تعداد زیادی ابزار نفوذ ایجادشده‌است که امکان ردیابی، نظارت، دستگیری و شناسایی ترافیک ناخواسته در یک شبکه یا یک دستگاه را فراهم‌می‌کند. در این ابزارهای متعدد نفوذ که در اوایل دهه ۲۰۰۰ معرفی‌شدند، عمده‌ترین ابزارهای مورداستفاده سیستم‌های پیشگیری از نفوذ و سیستم‌های تشخیص نفوذ هستند [۳].

استفاده گسترده از اینترنت منجر به افزایش تصاعدی حجم اطلاعات مبادله‌شده بین دستگاه‌های مختلف و تعداد روش‌های جدید حملات شبکه شده‌است. روش‌های مرسوم مانند فایروال که بر فیلتر کردن داده‌ها متمرکز است، ممکن است برای شناسایی به‌موقع انواع حملات مناسب نباشد. برای مدیریت مؤثر و شناسایی به‌موقع این نوع حملات، سیستم‌های تشخیص نفوذ مبتنی بر الگوریتم‌های یادگیری ماشینی برای پردازش کارآمد حجم زیادی از داده‌ها برای شناسایی هرگونه فعالیت مخرب بسیار مؤثر هستند. سیستم‌های تشخیص نفوذ مبتنی بر یادگیری ماشین برای تجزیه‌وتحلیل تمام فعالیت‌های شبکه برای هرگونه رفتار مخرب استفاده‌می‌شود [۴].

ازآنجاکه رشد سریع در استفاده از شبکه‌های کامپیوتری منجر به مسائل مربوط به حفظ در دسترس بودن، یکپارچگی و محرمانه بودن شبکه می‌شود، این امر مدیران شبکه را ملزم به اتخاذ انواع مختلفی از سیستم‌های تشخیص نفوذمی‌کند که به نظارت بر ترافیک شبکه برای فعالیت‌های غیرمجاز و مخرب کمک‌می‌کند. سیستم تشخیص نفوذ، ترافیک جاری در شبکه را از طریق سیستم‌های رایانه‌ای برای جستجوی فعالیت‌های مخرب و تهدیدات بررسی‌می‌کند و در صورت یافتن آن تهدیدها، هشدار ارسال‌می‌کند. برای این منظور، در این تحقیق، مدلی پیشنهادمی‌شود که برای حل مسئله تشخیص نفوذ در شبکه‌های کامپیوتری از تلفیق روش‌های XGBoost و خوشه‌بندی kMeans استفاده‌می‌کند. در روش پیشنهادی این تحقیق kMeans برای خوشه-بندی و اضافه کردن اطلاعات جدید به مجموعه داده استفاده‌خواهدشد. روش XGBoost برای شناسایی و طبقه‌بندی ترافیک به‌کار گرفته‌می-شود. برای بهبود عملکرد سیستم پارامترهای XGBoost تنظیم‌شده و ویژگی‌های موثر با استفاده از الگوریتم ژنتیک تعیین خواهدشد. در انتها تجزیه و تحلیل دقیق روش پیشنهادی و تعدادی از تکنیک‌های یادگیری ماشین انجام‌می‌شود و کارایی روش پیشنهادی مورد بررسی قرارمی‌گیرد.

بخش‌های باقی‌مانده از این کار به شرح زیر است: بخش ۲ کار مرتبط را توضیح‌می‌دهد. بخش۳ مدل ترکیبی پیشنهادی را مورد بحث قرارمی-دهد. جزئیات آزمایشی، بحث در مورد نتایج و به دنبال آن مقایسه با رویکردهای موجود در بخش ۴ ارائه‌شده‌است. در نهایت نتیجه‌گیری و کار های آتی در بخش‌های ۵ و ۶ ارائه‌شده‌است.

۲. پژوهش‌های مرتبط

با وجود پیشرفت‌های اخیر، سیستم‌های تشخیص نفوذ فعلی نیز مشکلاتی را در افزایش دقت تشخیص، افزایش سطوح هشدار کاذب و شناسایی فعالیت‌های مشکوک تجربه‌می‌کنند. به‌منظور پرداختن به این مسائل و ساختن یک سیستم تشخیص نفوذ مؤثر، محققان زیادی بر طراحی سیستم‌های تشخیص نفوذ متکی به رویکردهای یادگیری ماشین تمرکزکردند. در پروژه [۵] از الگوریتم سنجاقک برای انتخاب ویژگی و از الگوریتم جنگل تصادفی به منظور دسته‌بندی استفاده‌شد. داده‌های به‌کار رفته در پژوهش, مجموعه داده KDD-99 بود و عملیات متوازن-سازی در آن استفاده‌شد. مسئله با الگوریتم‌های مختلف یادگیری ماشین و یادگیری عمیق مورد آزمون قرارگرفت و بهترین الگوریتم انتخاب‌شد. در این تحقیق مقدار صحت در روش پیشنهادی مقدار ۹۹. ۸۳ به‌دست-آمد.

در تحقیق [۶] یک مدل ترکیبی IDS ارائه‌شد که به ادغام انتخاب ویژگی، طبقه‌بندی و بهینه‌سازی هایپرپارامترها پرداخته‌است. ابتدا، ویژگی‌های انبوه اولیه جداگانه و به روش‌های اطلاعات متقابل اصلاح‌شده (MMI)، الگوریتم ژنتیک (GA) و آزمون F تحلیل واریانس واردشدند و پس از آن اشتراک‌گیری از خروجی آن‌ها به عنوان ویژگی های نهایی مؤثر و کاهش‌یافته صورت‌پذیرفت. در ادامه، یک طبقه‌بند ترکیبی CNN و LSTM به‌کارگرفته‌شد که هایپرپارامترهای آن توسط یک الگوریتم بهینه‌سازی به نام گرگ خاکستری-نهنگ با جابه‌جایی تصادفی (RS-GWO-WOA) تعیین‌شد. این روش بر روی مجموعه داده NSL-KDD اعمال‌شد و به دقت ۷۶،۵۲٪ رسید .

در مقاله [۷] روشـــی با اســـتفاده از الگوریتم درخت تصـــمیم J48 ارائه شده‌است که می‌تواند نرخ تشخیص ترافیک نرمال و نرخ تشخیص حملات نوع DOS (Denial of Service attack) را در شبکه به بیش از ۹۹٪ برســـانند. در این تحقیق از مجموعه داده KDD CUP 99 که

مجموعه داده اســـتاندارد جهت آزمایش روش‌های تشـــخیص نفوذ به شبکه‌های کامپیوتری می‌باشد جهت آموزش و آزمایش استفاده‌شد. نرخ تشخیص ترافیک نرمال و حملات نوع DOS در این روش با استفاده از روش‌های کاهش ویژگی در مقایســـه با دیگر روش‌های مطرح‌شـــده، افزایش‌یافت.

در تحقیق [۸] مدلی پیشنهادمی‌گردد که الگوریتم‌های مختلف دسته‌بندی را روی مجموعه داده خود آزمایش‌کرده و نتایج شبیه‌سازی را نشان‌می‌دهد. بین تمامی الگوریتم‌ها، الگوریتم J48 دارای بالاترین مقدار درستی به میزان ۸۵/۴۹٪، دارای بالاترین میزان دقت به مقدار ۸۶/۵۷٪ و دارای بالاترین مقدار یادآوری به مقدار ۸۶/۵۷٪ می‌باشد. نوآوری اصلی در این تحقیق، استفاده از الگوریتم‌های مدل کاهل و مدل قانون محور است. در تحقیقات خارجی نیز فرگ و همکاران [۹] مروری بر الگوریتم‌های یادگیری عمیق سیستم‌های تشخیص نفوذ و مجموعه داده‌های تشخیص هرزنامه ارائه‌کردند و به این نتیجه رسیدند که مدل‌های یادگیری عمیق می‌توانند بهتر از مدل‌های یادگیری ماشین سنتی و متنی برای تشخیص نفوذ و هرزنامه عمل‌کنند.

در تحقیق [۱۰] بر اساس آزمایش‌ها و تحلیل‌های مقایسه‌ای انجام‌شده برای ویژگی‌های بهینه‌سازی ازدحام ذرات (PSO) و Xgboost، مدل PSO-Xgboost پیشنهادشد که با توجه به دقت طبقه‌بندی بالاتر آن نسبت به سایر مدل‌های جایگزین مانند Xgboost، Random Forest، Bagging و آدابوست نتایج مناسبی را داشته‌است. نتایج تجربی نشان‌داد که مدل PSO-Xgboost در دقت، یادآوری، میانگین کلان (ماکرو) و میانگین دقت متوسط، به‌ویژه هنگام شناسایی گروه‌های اقلیت حملات از سایر مدل‌های مقایسه‌ای بهتر عمل‌می‌کند. در مقاله [۱۱] یک مدل یادگیری ماشین مبتنی بر PCA برای طبقه‌بندی مجموعه داده‌های سیستم تشخیص نفوذ پیشنهادشد. نتایج تجربی این تحقیق، این واقعیت را تأییدمی‌کند که مدل پیشنهادی بهتر از مدل‌های یادگیری ماشین پایه موجود عمل‌می‌کند. دشموخ و همکاران [۱۲] حمله خاص مسیریابی را در یک محیط اینترنت اشیا گزارش کردند. این تحقیق بر روی شناسایی موقعیت نفوذگر و هشدار به مدیر شبکه متمرکز شد. شناسایی گره‌های همسایگی و استخراج تأثیر تهدید از دستاوردهای این پژوهش به شمار می‌رود. در تحقیق [۱۳] یک شبکه عصبی بازگشتی کانولوشن برای ساخت یک سیستم تشخیص نفوذ ترکیبی مبتنی بر یادگیری عمیق که حملات را از طریق یک شبکه شناسایی می‌کند، استفاده شد. برای افزایش کارایی سیستم تشخیص نفوذ و قابلیت پیش‌بینی، شبکه عصبی کانولوشن برای جمع‌آوری ویژگی‌های محلی به کار می رود در حالی که یک شبکه عصبی تکراری لایه‌ای عمیق ویژگی‌های سیستم تشخیص نفوذ شبکه هیبریدی مبتنی بر یادگیری عمیق (HDLNIDS) را استخراج می‌کند. آزمایش‌ها با استفاده از داده‌های معیار CICIDS-2018، برای تعیین اثربخشی سیستم پیشنهادی انجام‌شد و دقت متوسط ۹۸,۹۰ درصد در تشخیص حملات مخرب به‌دست‌آمد.

در تحقیق مانیموراگان و همکاران [۱۴] برای افزایش دقت و کاهش نرخ هشدار نادرست (FAR) در مسئله تشخیص نفوذ از الگوریتم بهینه‌سازی جستجوی کلاغ و سیستم استنتاج عصبی فازی تطبیقی (CSO-ANFIS) استفاده شد. نتایج آزمایش‌ها تشخیص نفوذ این تحقیق مناسب و کارآمدتر گزارش‌شد. در مقاله گلرنگ و همکاران [۱۵]، یک رویکرد ترکیبی چندهدفه برای شناسایی مؤثر حملات در شبکه پیشنهادشد. نتایج آزمایش‌ها نشان‌داد که چارچوب پیشنهادی در مقایسه با راه‌حل‌های موجود در ادبیات به نتایج بهتری دست‌یافت. در مقاله [۱۶]، یک سیستم تشخیص نفوذ مبتنی بر یادگیری عمیق برای تشخیص حمله ارائه‌شد. این مدل با مجموعه داده‌های ترافیک زمان واقعی CICIDS2018 و Edge_IIoT آموزش داده شد. عملکرد مدل هنگام آموزش و آزمایش با مجموعه داده‌ها، به ترتیب به میزان دقت ۱۰۰٪ و ۹۹,۶۴٪ دست‌یافت. در جدول۱ خلاصه اطلاعات مربوط به تعدادی از کارهای پیشین مرتبط آورده شده‌است. در این روش‌ها از مجموعه داده‌های متنوعی برای بررسی روش‌های پیشنهادی استفاده‌شده‌است.

در مقاله [۱۷]، یک سیستم تشخیص نفوذ دولایه مبتنی بر طبقه بند KNN برای جداســـازی ترافیک عادی از حمله و شـــبکه مصـــنوعی پرسپترون چندلایه برای تشخیص نوع حمله ارائه‌شد. در مجموعه داده KDD-CUP 99 نتایج آزمایش بیانگر دقت ۹۹,۷۴۳٪ بود.

در مقاله [۱۸]، از الگوریتم فرابتکاری رقابت امپریال (ICA) برای آموزش شبکه‌های عصبی استفاده و نشان داده‌شد که این الگوریتم در زمینه تشخیص نفوذ در سیستم اینترنت اشیا، می‌تواند عملکرد بسیار بهتری از منظر سرعت و دقت نسبت به روش‌های آموزشی کلاسیک داشـته‌باشـد. نتایج نشـان‌داد روش پیشـنهادی این تحقیق دارای دقت ۹۰٪ می‌باشد که در مقایسه با روش شبکه عصبی کلاسیک که دارای دقت ۷۵ درصد بوده عملکرد بهتری دارد.

جدول ۱: خلاصه اطلاعات تعدادی از روش‌های مرتبط

تحقیق	روش	مجموعه داده	ارزیابی
Mozafari, 2016	J48	DARPA	دقت، ۸۶,۵۷٪
Najafi Mohsen, 2015	درخت تصمیم J48	KDD CUP 99	دقت ۹۹٪
Mazloum, 2022	CNN-LSTM	NSL-KDD	دقت، ۷۶,۵۲٪
Niai, 2022	Dragonfly و جنگل تصادفی	KDD-9	صحت، ۸۳,۹۹
Ferrag et al, 2020	مدل CNN	CSE-CIC-IDS2018 Bot-IoT	نرخ تشخیص کلی ۹۷,۲۸٪ نرخ تشخیص کلی ۹۷,۰۱٪
Bhattacharya et al, 2020	XGBoost-Firefly	intrusion detection system (IDS)	دقت ۹۹,۹٪
Qazi, 2023	HDLNIDS	CICIDS-2018	دقت ۹۸,۹۰٪

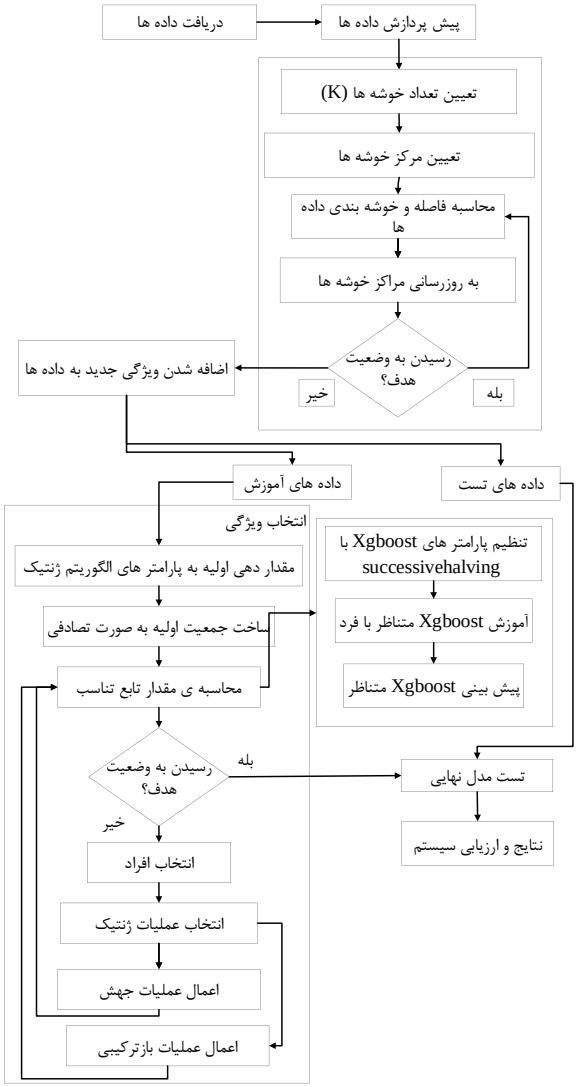
Hnamte, 2023	یادگیری عمیق	CICIDS2018 و Edge_IIoT	دقت ٪۹۹,۶۴
Kaffash et al, 2024	MLP-KNN	KDD-CUP 99	دقت ٪۹۹,۷۴۳
Nazarpour et al, 2023	NN-ICA	KDD-CUP	دقت ٪۹۰

۳. روش پیشنهادی

سیستم پیشنهادی این تحقیق از ترکیب الگوریتم‌های K-Means و XGboost برای تشخیص نفوذ استفاده‌می‌کند. در این تحقیق از روش Successive halving برای تنظیم پارامترها استفاده‌می‌شود. نحوه عملکرد سیستم پیشنهادی در شکل ۱ آمده‌است.

روش مدنظر در این پژوهش متشکل از الگوریتم‌های مختلف هوش مصنوعی می‌باشد. در ابتدا با دریافت داده، پیش‌پردازش لازم بر روی آن‌ها صورت می‌گیرد تا داده‌های موردبررسی برای استفاده در روش‌های یادگیری ماشین آماده‌شوند. در ادامه با اعمال روش خوشه‌بندی K-Means ، اقدام به خوشه‌بندی داده‌ها می‌شود. در این بخش برچسب‌های مجموعه داده ناد یده گرفته‌می‌شود و از آن در روش K-Means استفاده‌نمی‌شود. به این ترتیب درحقیقت مجموعه داده‌ها به عنوان داده‌های بدون برچسب در نظر گرفته‌می شود. پس از اعمال روش K-Means، شماره مربوط به هریک از خوشه‌های حاصل به‌عنوان یک ویژگی اضافه به مجموعه داده مورد بررسی اضافه می‌شود.

با توجه به اینکه از روش دسته‌بندی XGBoost در گام بعدی استفاده‌می‌شود، ابتدا با روش نصف کردن پی‌درپی، پارامترهای مناسب برای روش XGBoost تعیین‌می‌شود. در ادامه به‌منظور اعمال فرآیند انتخاب ویژگی از ویژگی های مجموعه داده موردنظر، روش ژنتیک اعمال‌می‌شود و هر بار، جایگشت متفاوتی از زیرمجموعه ویژگی‌های مجموعه داده موردبررسی، ارزیابی‌می‌شود و بهترین نتایج حاصل گزارش‌می‌شود.



شکل ۱: عملکرد سیستم پیشنهادی

۱،۳. الگوریتم XGBoost

در معادله (۱) که تابع هدف XGBoost را نشان‌می‌دهد، عبارت $L(y_i,\hat{y}_i)$ تابع ضرر است که در آن y_i ارزش واقعی و $\hat{y}_i$ارزش پیش‌بینی مدل است. تابع ضرر می‌تواند مربع انحرافات $\hat{y}_i$ پیش‌بینی از y_i واقعی باشد. $\Omega\ (f_i\ )$ نیز یک عبارت منظم‌سازی است برای جریمه کردن پیچیدگی مدل جهت جلوگیری از برازش بیش‌ازحد (over fitting).

$$Obj\ (t) = \sum_{i=1}^{n} L\ (y_i, \hat{y}_i)\ + \sum_{i=1}^{n} \Omega\ (f_i) \qquad (۱)$$

در معادله (۲) که عبارت منظم‌سازی را نمایش‌می‌دهد، γ ضریب جریمه اندازه درخت است که در تعداد برگ‌های درخت یعنی T ضرب می‌شود و اندازه درخت را کنترل‌می‌کند λ نیز ضریب جریمه وزن برگ است که در نرم دو (L2) وزن برگ‌ها ضرب می شود.w نیز وزنی است که به هر برگ اختصاص‌دادیم.

$$\Omega\ (f_t) = \gamma T + \frac{1}{2}\lambda \sum_{i=1}^{T} w_j^2 \qquad (۲)$$

بعد از t تکرار، مدل پیش‌بینی به‌صورت معادله (۳) نوشته‌می‌شود:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \qquad (۳)$$

بنابراین تابع هدف به شکل معادله (۴) درمی‌آید:

$$Obj\ (t) = \sum_{i=1}^{n} \mathrm{L}\ (y_i, \hat{y}_i^{(t-1)} + f_t(x_i))\ + \Omega\ (f_t) \qquad (۴)$$

XGBoost متفاوت با GBDT[۱] است. XGBoost می‌تواند تابع ضرر را بر اساس مشتق دوم فرمول تیلور گسترش دهد به‌طوری‌که تابع هدف جدید به‌دست‌آمده نرخ همگرایی سریع‌تر و دقت بالاتری نسبت به GBDT اصلی داشته‌باشد. بر این اساس تابع هدف نهایی به شکل معادله (۵) نوشته‌می‌شود:

$$Obj\ (t) \cong \sum_{i=1}^{n}[\mathrm{L}\ (y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)]\ + \Omega\ (f_t) \qquad (۵)$$

$$g_i = \frac{\partial\ (\mathrm{L}\ (y_i, \hat{y}_i^{(t-1)}))}{\partial\ \hat{y}_i^{(t-1)}} \qquad (۶)$$

$$h_i = \frac{\partial^2\ (\mathrm{L}\ (y_i, \hat{y}_i^{(t-1)}))}{\partial\ \hat{y}_i^{(t-1)}} \qquad (۷)$$

که در آن g_i و h_i همان‌طور که در معادله‌های (۶) و (۷) داده‌شده به ترتیب مشتقات اول و دوم تابع ضرر $L(\phi)$ هستند.

۲،۳. الگوریتم k-Means

یکی از روش‌های معروفی که در دیدگاه خوشه‌بندی مطرح می‌باشد، الگوریتم k-Means است. این روش با مقداری مشخص برای (k تعداد دسته‌ها) شروع و سعی می‌کند که مجموعه‌ای از نمونه‌های مشخص را در این K گروه دسته‌بندی کند طوری‌که فرضیات موردنظر، رعایت شوند. لازم به ذکر است که معیار تخصیص هر نمونه به یک کلاس نیز در این روش حداقل مسافت اقلیدسی هر نمونه از نقطه میانی (مرکزی یا نماینده) هر کلاس یا دسته است. مراحل مهمی که در این الگوریتم طی می‌گردد به‌طور خلاصه به شرح ذیل می‌باشد [۱۷].

۱. تشکیل K کلاس و یا دسته به صورت $\{C1, C2, \ldots Ck\}$ برای خوشه‌بندی m نمونه از مجموعه M.

۲. محاسبه بردار z_j بر اساس رابطه زیر که نشان‌دهنده مرکز یا نماینده هر دسته c_j می‌باشد.

$$z_j = \frac{\sum_{x \in c_j} x}{\#c_j} \quad for\ j = 1,2, \ldots, K \qquad (۸)$$

در این رابطه، x نشان‌دهنده بردار نمونه‌ای است که در Cj عضو می‌باشد و #Cj نشان‌دهنده تعداد j عضو می‌باشند. لازم به ذکر است که رابطه بالا برای محاسبه مرکز نمونه‌هایی می‌باشد که در کلاس C هر دسته در خلال حل استفاده‌می‌شود و در شروع الگوریتم معمولاً K نمونه تصادفی انتخاب و به‌عنوان مرکز هر دسته در نظر گرفته‌می‌شوند

۳. محاسبه تابع هدف ناشی از دسته‌بند $\{C1, C2, \ldots Ck\}$ بر اساس رابطه زیر که به محاسبه مجموع فواصل نمونه‌ها از مرکز دسته‌ها می‌پردازد.

$$f(C_1, C_2, \ldots, C_K) = \sum_{j=1}^{k} \sum_{x \in c_j} |x - z_j|^2 \qquad (۹)$$

۴. کمینه نمودن تابع هدف رابطه بالا و یافتن دسته‌بندی مناسب روی مجموعه M با تعداد دسته K.

۳،۳. نحوه انتخاب پارامترها

در این تحقیق از روش Successive halving یعنی نصف کردن پی‌درپی برای جستجوی فضای پارامتر برای بهینه‌سازی پارامترهای مدل یادگیری استفاده‌شده‌است.. استراتژی جستجو در اولین تکرار، شروع به ارزیابی همه نامزدها (ترکیب پارامترها) با مقدار کمی از منابع (تعداد نمونه آموزشی) می‌کند. تعداد محدودی از بهترین نامزدها برای تکرار بعدی انتخاب‌می‌شوند و منابع بیشتری به آن‌ها اختصاص داده‌می‌شود. فقط زیرمجموعه‌ای از نامزدها تا آخرین تکرار زنده‌می‌مانند. این‌ها نامزدهایی هستند که دائماً در بین نامزدهای دارای رتبه برتر در همه تکرارها رتبه‌بندی می‌شوند.

۴،۳. انتخاب ویژگی

در این پژوهش از روش ژنتیک برای انتخاب زیرمجموعه‌ای از داده‌ها استفاده‌می‌شود. به کمک روش ژنتیک می‌توان زیرمجموعه‌های مختلف مجموعه داده موردنظر را بررسی و بهترین نتایج را گزارش‌کرد. به‌این‌ترتیب در الگوریتم ژنتیک سیستم پیشنهادی هر راه‌حل یا فرد نشان‌دهنده تعدادی ویژگی است که برای به‌دست‌آوردن میزان مطلوبیت آن مدل Xgboost برای ویژگی‌های متناظر ایجادمی‌شود و دقت مدل پیشنهادی به‌عنوان میزان تابع تناسب برای فرد موردنظر ثبت‌می‌شود. درنهایت کروموزوم با بیشترین تناسب به‌عنوان بهترین راه‌حل شناخته‌می‌شود. این کروموزوم بهترین ویژگی‌های ممکن را شناسایی کرده‌است و مدلی که با استفاده از این ویژگی‌ها ساخته‌شود بالاترین کارایی را ارائه‌خواهدداد. درواقع در الگوریتم ژنتیک پیشنهاد شده هدف یافتن کروموزومی است که بهترین زیر مجموعه از ویژگی ها را نشان‌می‌دهد و تناسب هر کروموزوم با بررسی دقت پیاده‌سازی روش Xgboost با درنظرگرفتن ویژگی‌های متناظر با کروموزوم مربوطه به دست می‌آید.

۵،۳. روش‌های ارزیابی سیستم پیشنهادی

در ادامه با استفاده از روش ارزیابی صحت[۲] و بازخوانی[۳] داده‌های دسته‌بندی‌شده و دقت[۴] نهایی دسته‌بندی انجام‌شده الگوریتم مورد ارزیابی قرارمی‌گیرد [۲]. این معیارها به‌صورت زیر تعریف‌می‌شوند:

دقت برابر است با تعداد اسناد بازیابی شده مرتبط تقسیم‌بر تعداد کل اسناد بازیابی‌شده.

بازخوانی برابر است با تعداد اسناد بازیابی‌شده مرتبط تقسیم‌بر تعداد کل اسناد مرتبط واقعی در پایگاه داده.

برای بررسی بهتر سیستم پیشنهادی و ویژگی‌هایی که این سیستم دارد نتایج به‌دست‌آمده برای معیارهای ارزیابی مختلف با نتایج به‌دست‌آمده از روش‌های پایه مقایسه‌می‌شود و مزایا و معایب آن‌ها مشخص‌می‌شود.

۶،۳. مجموعه داده موردبررسی

مجموعه داده‌های مورداستفاده در این مطالعه NSL-KDD است که در بسیاری از مطالعات تشخیص نفوذ استفاده‌می‌شود. این پایگاه داده حاوی حجم زیادی از داده‌های مربوط به حملات مختلف است که در محیط شبیه سازی شده یک شبکه کامپیوتری در نیروی هوایی آمریکا جمع‌آوری‌شده است. این داده ها حاصل ۹ هفته تراکنش های به‌د ست‌آمده در این شبکه ا ست. داده‌ها در مجموعه داده NSL-KDD به‌عنوان عادی یا به‌عنوان یکی از ۲۴ نوع مختلف حمله برچسب‌گذاری می‌شوند. این ۲۴ حمله را می‌توان به چهار کلاس طبقه‌بندی کرد: Probe، DoS، R2L و U2R

در حملات DoS (Denial of Service) تلاشی مخرب برای مسدود کردن منابع و سرویس‌های یک سیستم یا شبکه صورت‌می‌گیرد.

در حملات نوع Probe، اطلاعات مربوط به آسیب‌پذیری‌های احتمالی سیستم هدف جمع‌آوری‌می‌شود و بعداً برای حمله به سیستم‌ها استفاده-می‌شود.

در حملات R2L (Remote to Local)، توانایی غیرمجاز برای ریختن بسته‌های داده به سیستم راه دور از طریق شبکه و د ستر سی به‌عنوان کاربر یا روت برای انجام فعالیت‌های غیرمجاز وجوددارد.

در حملات U2R (User to Root)، مهاجمان به‌عنوان یک کاربر عادی به سیستم دسترسی پیدامی‌کنند و آسیب‌پذیری‌ها را برای به‌دست آوردن امتیازات مدیریتی می‌شکنند.

محیط پیاده‌سازی داده‌ها محیط برنامه‌نویسی پایتون بود.

۷،۳. آماده‌سازی داده‌ها

آماده‌سازی داده‌ها فرآیند پاک‌سازی و تبدیل داده‌های خام قبل از پردازش و تجزیه‌وتحلیل است. این‌یک مرحله مهم قبل از پردازش است و اغلب شامل قالب‌بندی مجدد داده‌ها، انجام اصلاحات در داده‌ها و ترکیب مجموعه داده‌ها برای غنی‌سازی داده‌ها است. آماده‌سازی داده‌ها با انجام تبدیل ویژگی با کدگذاری One-hot، نرمال‌سازی با روش Z-score و محدود کردن حالات حمله و نرمال به ۵ مورد انجام‌شد. داده‌ها در یک کلاس عادی و چهار کلاس حمله طبقه‌بندی‌می‌شوند. چهار کلاس حمله با عنوان‌های DoS، Probe، R2L و U2R گروه‌بندی می‌شوند.

۴. نتایج

برای بررسی بهتر نتایج حاصل از سیستم پیشنهادی، ابتدا به بیان نتایج حاصل از اعمال مدل‌های مختلف بر روی مجموعه داده خواهیم‌پرداخت. مدل‌های موردبررسی در این پژوهش، شامل روش‌های درخت تصمیم، XGBoost، KNN و جنگل تصادفی می‌باشد. معیارهای مختلفی برای مقایسه این روش‌ها موردبررسی قرارگرفته‌اند که شامل Precisoin، Recall و F1-Score می‌باشند.

۱،۴. نتایج اعمال الگوریتم‌های مختلف

در این بخش نتایج به‌دست‌آ مده از اعمال روش درخت تصمیم، XGBoost، Knn، بر روی مجموعه داده NSL-KDD نشان داده‌شده است. جدول ۲ نتایج به‌دست‌آمده از اعمال روش درخت تصمیم، kNN،XGBoost، در معیارهای ارزیابی Precisoin، Recall و F1-Score را به ترتیب در کلاس‌های Dos، Normal، Probe، R2L و U2R نشان‌می‌دهد.

جدول ۲: نتایج به‌دست‌آمده از اعمال روش‌های مختلف

درخت تصمیم	precision	recall	f1-score	
Dos	94.74%	78.73%	86.00%	
Normal	69.61%	93.42%	79.78%	
Probe	57.99%	66.25%	61.85%	
R2L	84.82%	11.55%	20.34%	
U2R	58.06%	26.87%	36.73%	
accuracy				75.49%
XGBoost				
Dos	96.25%	74.71%	84.13%	
Normal	73.74%	95.27%	83.13%	
Probe	76.28%	70.01%	73.01%	
R2L	13.79%	7.53%	9.74%	
U2R	4.88%	26.87%	8.26%	
accuracy				74.84%
kNN				
Dos	94.55%	79.58%	86.43%	
Normal	66.89%	92.92%	77.78%	
Probe	70.93%	69.64%	70.28%	
R2L	80.20%	5.83%	10.87%	
U2R	38.46%	29.85%	33.61%	
accuracy				75.25%
جنگل تصادفی				
Dos	97.14%	68.95%	80.65%	
Normal	60.52%	98.99%	75.12%	
Probe	92.18%	47.25%	62.48%	
R2L	0.00%	0.00%	0.00%	
U2R	0.00%	0.00%	0.00%	
accuracy				71.06%

طبق نتایج جدول ۲، میزان دقت (Accuracy) کلی به‌دست‌آمده از اعمال در خت تصمیم بر روی این مجموعه داده برابر با ۷۵/۴۹٪ می‌باشد. میزان دقت کلی حاصل از اعمال XGBoost بر روی این مجموعه داده برابر با ۷۴/۸۴٪ می‌باشد. به‌این‌ترتیب می‌توان دریافت که در مورد این مجموعه داده دقت کلی روش درخت تصمیم از XGBoost بیشتر است. میزان دقت کلی با اعمال جنگل تصادفی بر روی این مجموعه داده برابر با ۷۱/۰۶٪ می‌باشد که بدترین میان دقت در بین تمامی روش‌های برر سی شده ا ست. میزان دقت کلی kNN بر روی این مجموعه داده برابر با ۷۵/۲۵٪ می‌باشد.

ا نتایج ثبت‌شده در جدول می‌توان در یافت روش های در خت تصمیم، XGBoost و kNN عملکرد خوبی از خود نشان‌داده‌اند. در فرآی ند توسعه روش پیشنهادی عملکرد این روش ها در بخش طبقه‌بندی سیستم مورد بررسی قرارگرفت و روش XGBoost بیشترین بهبود را با تنظیم پارامتر ها به‌دست‌آورد و به عنوان روش یادگیری در سیستم پیشنهادی به کار گرفته‌شد.

۲،۴. پارامترهای تعیین‌شده به کمک روش Successive halving

این تحقیق برای بهینه‌سازی XGBoost از روش Successive halving بهره‌برده‌است. پارامترها نیز n_estimators، min_samples_split، min_samples_leaf و max_depth می‌باشند.

n_estimators: تعداد مراحل تقویتی برای انجام اســـت. تعداد زیاد این عملگر معمولاً عملکرد بهتری را به همراه دارد. مقادیر باید در محدوده [۱، inf) باشد.

min_samples_split: یعنی حداقل تعداد نمونه موردنیاز برای تقسیم یک گره داخلی. اگر int، مقادیر باید در محدوده [۲، inf) باشد.

min_samples_leaf: حداقل تعداد نمونه موردنیاز برای قرارگرفتن در یک گره برگ است. یک نقطه تقسیم در هر عمق تنها در صورتی در نظر گرفته‌می‌شـود که حداقل نمونه‌های آموزشـی min_samples_leaf در هر یک از شاخه‌های چپ و راست باقی‌بماند. این مورد ممکن است اثر هموار سازی مدل را به‌خصوص در رگرسیون داشته‌باشد. اگر int، مقادیر باید در محدوده [۱، inf) باشد.

max_depth: حداکثر عمق برآوردگرهای رگرسـیون فردی. حداکثر عمق تعداد گره‌ها را در درخت محدود می‌کند. این پارامتر باید برای بهترین عملکرد تنظیم‌شـــود. بهترین مقدار به تعامل متغیرهای ورودی بستگی دارد. مقادیر باید در محدوده [۱، inf) باشد.

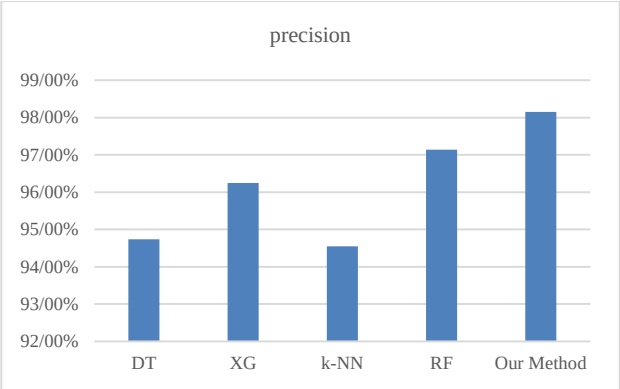
در این تحقیق، مقادیر به‌دست‌آمده به کمک روش successive halving، برای پارامترهای n_estimators، min_samples_split، min_samples_leaf و max_depth به ترتیب برابر با ۱۵۶، ۲، ۱ و ۴ می‌باشد. (جدول ۳)

جدول ۳: نتایج اعمال روش پیشنهادی

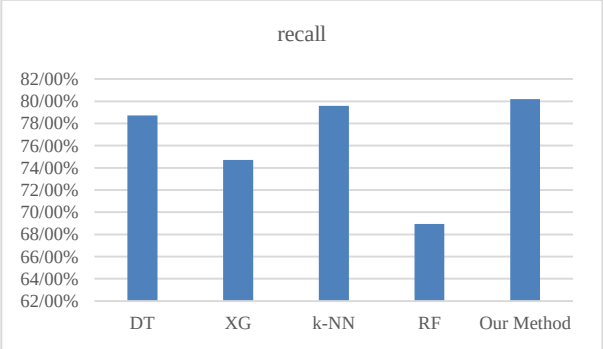
	Precision	recall	f1-score	
Dos	98.15%	80.18%	87.37%	
Normal	76.24%	96.47%	85.24%	
Probe	81.19%	73.43%	76.21%	
R2L	86.35%	26.72%	29.57%	
U2R	65.13%	33.84%	41.68%	
accuracy				78.92%

طبق جدول ۳، روش پیشـنهادی در این پژوهش، نسـبت به سـایر راهکارهای موردبررسـی، دقت بالاتری برای شـناسـایی حالات حمله و نرمال ارائه‌می‌دهد. برای تصویر سازی بهتر نتایج حاصل از پیاده سازی، نمودار میله‌ای برای حمله DoS با توجه به معیارهای مختلف در ادامه نشان داده‌شده‌است.

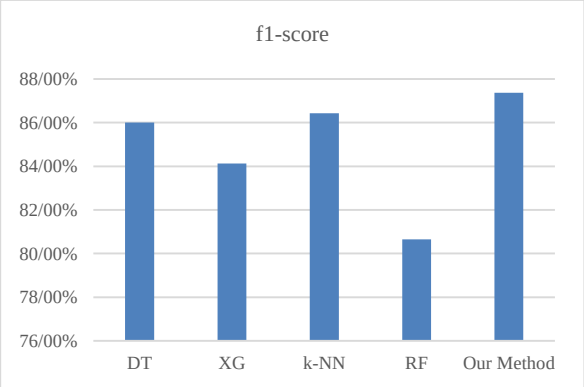
مقایسه روش پیشنهادی با سایر راهکارهای موردبررسی برای شناسایی حمله DoS با توجه به معیار Precision در شکل ۲ نشان داده‌شده است.



شکل ۲: مقایسه روش پیشنهادی با سایر راهکارها در شناسایی حمله DoS با توجه به معیار Precision

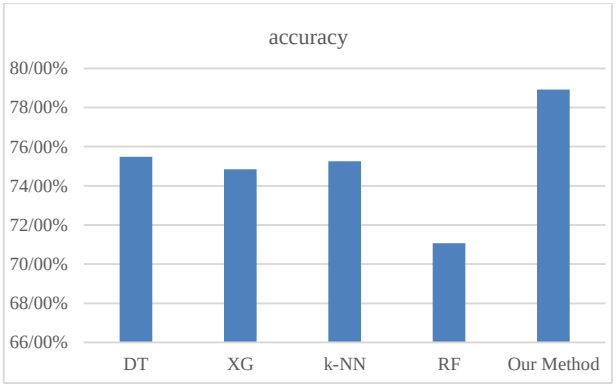


شکل ۳: مقایسه روش پیشنهادی با سایر راهکارهای موردبررسی برای شناسایی حمله DoS با توجه به معیار Recall



شکل ۴: مقایسه روش پیشنهادی با سایر راهکارهای موردبررسی برای شناسایی حمله DoS با توجه به معیار F1-Score

مقایسه روش پیشنهادی با سایر راهکارهای موردبررسی برای شناسایی حمله DoS با توجه به معیار Accuracy در شکل ۵ نشان داده‌شده است.



شکل ۵: مقایسه روش پیشنهادی با سایر راهکارهای موردبررسی برای شناسایی حمله DoS با توجه به معیار Accuracy

در جدول ۴، نتایج مقایسه روش مطرح‌شده در این پژوهش و راهکار پژوهش [۱۸] از منظر معیار F-Measure ارائه‌شده‌است. این تحقیق نیز نتایج خود را روی چهار کلاس مختلف مجموعه داده NSL-KDD جداگانه ثبت‌کرده‌است.

مقایسه نشان‌می‌دهد راهکار پیشنهادی تحقیق حاضر در زمینه حملات DOS، U2R و Normal عملکرد بهتری دارد.

جدول ۴: مقایسه معیار F-Measure در سیستم پیشنهادی و روش ارائه‌شده در مقاله [۱۸]

دسته رکورد	روش ارائه‌شده مقاله [۱۸]	سیستم پیشنهادی
تست DoS	77.19%	87.37%
تست U2R	14.17%	41.68%
تست R2L	39%	29.57%
تست Probe	76.40%	76.21%
تست Normal	75.33%	85.24%

۵. نتیجه‌گیری

سیستم‌های تشخیص نفوذ، وظیفه شناسایی و تشخیص هرگونه استفاده غیرمجاز از سیستم و سوءاستفاده یا آسیب‌رسانی توسط هر دودسته کاربران داخلی و خارجی را بر عهده‌دارند. تشخیص و جلوگیری از نفوذ امروزه به‌عنوان یکی از مکانیزم‌های اصلی در برآوردن امنیت شبکه‌ها و سیستم‌های رایانه‌ای مطرح است و عموماً در کنار دیواره‌های آتش و به‌صورت مکمل امنیتی برای آن‌ها مورداستفاده قرارمی‌گیرند. سیستم‌های تشخیص نفوذ به‌صورت سامانه‌های نرم‌افزاری و سخت‌افزاری ایجادشده‌اند و هرکدام مزایا و معایب خاص خود را دارند. سرعت و دقت از مزایای سیستم‌های سخت‌افزاری است و عدم شکست امنیتی آن‌ها توسط نفوذ گران، قابلیت دیگر این‌گونه سیستم‌ها می‌باشد؛ اما استفاده آسان از نرم‌افزار، قابلیت سازگاری در شرایط نرم‌افزاری و تفاوت سیستم‌های عامل مختلف، عمومیت بیشتری را به سامانه‌های نرم‌افزاری می‌دهد و عموماً این‌گونه سیستم‌ها انتخاب مناسب‌تری هستند. با توجه به اهمیت این موضوع در این پژوهش به ارائه راهکاری مبتنی بر تلفیق روش‌های XGBoost و خوشه‌بندی kMeans برای تشخیص نفوذ در شبکه‌های کامپیوتری پرداخته‌شد.

طبق پیاده سازی صورت‌گرفته و نتایج حاصل، اعمال روش تلفیقی XGBoost و خوشه‌بندی kMeans با درنظرگرفتن ملاحظات مطرح‌شده، منجر به دستیابی به نتایج مناسبی می‌شود. همچنین بهره‌گیری از روش ژنتیک به شناسایی ویژگی های مهم‌تر در مجموعه داده کمک‌می‌کند.

در این آزمایش سیستم پیشنهادی و روش مطرح شده در پژوهش [۱۸] مقایسه‌شده‌اند. اقدم و همکاران در سال ۲۰۱۶ با بهره‌گیری از الگوریتم کلونی مورچه، راهکاری برای تعیین بهترین ویژگی‌ها در مسئله نفوذ به سیستم‌های کامپیوتری ارائه‌داده‌اند. برای ارزیابی روش مطرح‌شده در پژوهش [۱۸] از مجموعه داده NSL-KDD بهره برده شده‌است.

۶. کارهای آتی

در این تحقیق راهکاری با استفاده از ترکیب روش‌های XGBoost، KMeans و الگوریتم ژنتیک برای حل مسئله تشخیص نفوذ ارائه‌شده است. بااین‌حال با توجه به اهمیت و ویژگی‌های خاص این مسئله، پیدا کردن روش‌های به‌روز پیوسته موردنیاز است. ازآنجاکه با تمرکز بر ویژگی‌های مهم، عملکرد پیش‌بینی و استحکام و قابلیت تفسیر مدل افزایش می‌یابد، توجه به این بخش و به‌کاربردن سایر روش‌های موجود اهمیت زیادی دارد. برای ادامه این پژوهش می‌توان از سایر روش‌های بهینه‌سازی به‌جای روش ژنتیک بهره‌برد. روش‌های بهینه‌سازی تکاملی متعددی ارائه‌شده‌است ازجمله الگوریتم بهینه‌سازی ازدحام ذرات، جستجوی هارمونی، گرگ خاکستری، وال، ممتیک و ....

از تکنیک های یادگیری تجمعی نیز می‌توان برای بهبود نتایج کمک‌گرفت. ازجمله روش‌های تجمعی شناخته‌شده، روش بگینگ و پشته سازی می‌باشد که امکان بهبود عملکرد دسته‌بندهای ضعیف‌تر را فراهم‌می‌کند.

در این پژوهش به دلیل در اختیار نبودن مجموعه داده‌های داخلی، از داده‌های موجود در وب استفاده‌شده‌است. برای رفع این محدودیت، پژوهشگران می‌توانند با بررسی بیشتر، تأمین‌کنندگان داده‌ها در حوزه امنیت داخل کشور را شناسایی‌نموده و در صورت امکان از راهکار پیشنهادی بر روی‌داده‌های داخلی بهره‌ببرند. علاوه بر این، توسعه ارزیابی‌های فعلی با استفاده از ترافیک شبکه اصلی برای نشان‌دادن ارزش مدل پیشنهادی یکی از مواردی است که در ادامه می توان به آن پرداخت.

References

[1] H. N. Mohsenabad, A. Asghari, “Intrusion detection in computer networks using data mining techniques based on feature selection,” The first international conference on new research achievements in electrical and computer

engineering, Amirkabir University, Tehran, 2016 https://civilica.com/doc/496737 [Persian]
[2] T. Fawcett, "An introduction to ROC analysis," Pattern recognition letters, vol. 27, no. 8, pp. 861-74, 2006.
[3] S. Thapa, M. Akalanka, "The role of intrusion detection/prevention systems in modern computer networks: A review," In Conference: Midwest Instruction and Computing Symposium (MICS), vol. 53, pp. 1-14. 2020.
[4] M.K. Yadav, K.P. Sharma, "Intrusion Detection System using Machine Learning Algorithms: A Comparative Study," In 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), pp. 415-420. IEEE, 2021.
[5] M. Niai, J. Tannah, G. Shahmohammadi, A. Pourabrahimi, A Model for Multi-Class Intrusion Detection Using Machine Learning and Dragonfly Feature Selection, Scientific Journal of Electronic & Cyber Defense, Vol. 10, No. 3, Autumn 2022, Serial No. 39.
[6] J. Mazloum, H. Begdali, Integrated optimal hybrid deep neural network with feature selection for intrusion detection system in cyber attacks, electronic and cyber defense, 2022 [Persian]
[7] H. Najafi Mohsen Abad, M. Ketabi, intrusion detection in computer networks using data mining techniques, the third international conference on recent innovations in electrical and computer engineering, Tehran, 2015 [Persian] https://civilica.com/doc/576765
[8] M. Mozafari, Development of an intrusion detection system (IDS) in computer networks using Bayesian networks, 2016. [Persian]
[9] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, H. Janicke, "Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study," Journal of Information Security and Applications, vol. 50, Article ID 102419, 2020.
[10] J. Hui, Z. He, G. Ye, and H. Zhang, "Network intrusion detection based on PSO-XGBoost model," IEEE Access vol. 8, pp. 58392-58401, 2020.
[11] S. Bhattacharya, P. K. R. Maddikunta, R. Kaluri, S. Singh, T. R. Gadekallu, M. Alazab, U. Tariq, "A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU," Electronics, vol. 9, no. 2, p. 219, 2020.
[12] S. Deshmukh-Bhosale, S.S. Sonavane, "A real-time intrusion detection system for wormhole attack in the RPL based Internet of Things," Procedia Manufacturing, vol. 32, pp. 840-847, 2019.
[13] E.U.H. Qazi, M.H. Faheem, T. Zia, "HDLNIDS: hybrid deep-learning-based network intrusion detection system" Applied Sciences, 13(8), p.4921, 2023.
[14] S. Manimurugan, M. Al-qdah, M. Mustaffa, C. Narmatha, R. Varatharajan, "Intrusion detection in networks using crow search optimization algorithm with adaptive neuro-fuzzy inference system," Microprocessors and Microsystems, vol. 79, p. 103261, 2020.
[15] A. Golrang, A. Mohammadi Golrang, S. Yildirim Yayilgan, O. Elezaj, "A novel hybrid IDS based on modified NSGAII-ANN and random forest," Electronics, vol. 9, no. 4, p. 577, 2020.
[16] V. Hnamte, H. Jamal, "DCNNBiLSTM: An efficient hybrid deep learning-based intrusion detection system", Telematics and Informatics Reports 10 (2023): 100053.
[17] A. Kaffash, S.R. Kamel, M. Kheirabadi, "An Effective and Lightweight Intrusion Detection for IoT based on Fog and Cloud using KNN Classification," intelligent multimedia processing and communication systems(IMPCS), 5 (2), 55-64, 2024.
[18] M. Nazarpour, N. Nezafati, S. Shokouhyar, "Using the Modified Colonial Competition Algorithm to Increase the Speed and Accuracy of the Intelligent Intrusion Detection System," intelligent multimedia processing and communication systems(IMPCS), 4 (1), 1-10, 2023.
[19] S. S. Ghanadpour, A. Hezarkhani, "Investigating the behavior of copper element with respect to molybdenum, lead and zinc elements in Perkam porphyry copper deposit in Kerman province, using K-Means method," Advanced applied geology, vol. 3, No. 1, pp. 43-53, 2013. [Persian]
[20] M. H. Aghdam, P. Kabiri, "Feature selection for intrusion detection system using ant colony optimization," International Journal of Network Security, Vol.18, no.3, PP.420-432, 2016

پی‌نوشت

[1] Gradian Boosting Dicision Tree
[2] Precision
[3] Recall
[4] Accuracy