Journal of New Achievements in Electrical, Computer and Technology
Vol 2, No 1, Mar-Jun 2022,(92-106)
WWW.JNAECT.IR

مجله دستاوردهای نوین در برق،کامپیوتر و فناوری
دوره ۲، شماره ۱، بهار ۱۴۰۱ (۱۰۶-۹۲)

بهبود الگوریتم ماشین بردارپشتیبان جهت افزایش دقت تشخیص نفوذ توسط الگوریتم خفاش

مهشید صالحی[۱]*،سید رضا کامل[۲] ،حسن شاکری [۳]

[۱] دپارتمان مهندسی کامپیوتر ،واحد مشهد، دانشگاه ازاد اسلامی ، مشهد،ایران salehy.md@gmail.com

[2] دپارتمان مهندسی کامپیوتر ،واحد مشهد، دانشگاه ازاد اسلامی ، مشهد،ایران rezakamel@ieee.org

[3] دپارتمان مهندسی کامپیوتر ،واحد مشهد، دانشگاه ازاد اسلامی ، مشهد،ایران shakeri@mshdiau.ac.ir

چکیده :

سیستم تشخیص نفوذ دستگاه یا برنامه نرم افزاری است که شبکه یا سیستم ها را از نظر فعالیت های مخرب یا نقض خط مشی ها کنترل می کند. یکی از چالش های مهم در این زمینه، تشخیص درست حالت نرمال و حمله در سیستم میباشد. پژوهش های بسیاری در زمینه سیستم های تشخیص نفوذ میتنی بر روش های یادگیری صورت گرفته است . دقت سیستم تشخیص نفوذ اخیراً از طریق مدل‌های مختلف یادگیری ماشین بهبود یافته است. با این حال، دقت سیستم‌های تشخیص نفوذ همچنان یک چالش باقی می‌ماند، زیرا مهاجمان مرتباً رفتار خود را تغییر می‌دهند. این تحقیق مدلی را برای افزایش دقت سیستم تشخیص نفوذ با استفاده از ماشین بردار پشتیبان و الگوریتم خفاش پیشنهاد می‌کند.ماشین بردار پشتیبان یکی از الگوریتم های یادگیری ماشین است که توسط محققان بسیاری مورد تحلیل و بررسی قرار گرفته است. ماشین بردار پشتیبان در کار با حجم بالای داده ها کارایی لازم را دارد . به هر حال، عملکرد ماشین بردار پشتیبان به شدت به پارامترهای آن بستگی دارد. مقادیر مختلف پارامتر های ماشین بردار پشتیبان نتایج مختلفی را ارائه میدهند. بنابراین انتخاب ویژگی توسط الگوریتم خفاش انجام و پارامتر های ماشین بردار پشتیبان توسط الگوریتم خفاش بهبود پیدا میکند تا درصد دقت در سیستم های تشخیص نفوذ افزایش یابد. برای بررسی عملکرد مدل پیشنهادی از مجموعه داده Nsl-Kdd استفاده شده است . مدل پیشنهادی به کمک نرم افزار Matlab پیاده سازی شده است. نتایج نهایی نشان میدهد که روش انتخاب ویژگی با الگوریتم خفاش و بهبود الگوریتم ماشین بردارپشتیبان توسط الگوریتم خفاش در سیستم های تشخیص نفوذ برابر ۹۷.۸۶ درصد میباشد که نشان دهنده ۵.۴۲ درصد بهبود نسبت به مدل پیشنهادی یکی از اخرین کار های انجام شده که با ان مقایسه شده است ، می باشد.

کلمات کلیدی : سیستم تشخیص نفوذ ، K- نزدیکترین همسایه،ماشین بردار پشتیبان ، الگوریتم بهینه سازی خفاش ، متلب ، پایگاه داده Nsl-Kdd .

۱.مقدمه

در سال‌های گذشته در صنایع، مشاغل مختلف و...زمینه های زندگی بشری به طور گسترده از شبکه های کامپیوتری استفاده می شود(Almseidin et al., n.d.). همانطور که اینترنت به سرعت رشد می کند، تعداد تهاجمات شبکه نیز به طور قابل توجهی افزایش میابد(Yao et al., 2018). دسترسی فوری به شبکه جهانی اینترنت، مردم و سازمان ها را در معرض تهدیدات سایبری قرار می دهد پس نرم افزارهای آنتی ویروس مختلفی برای محافظت از حریم خصوصی کاربران و داده های مهم استفاده میشود(Drewek-Ossowicka et al., 2020). فایروال ها و سیستم های تشخیص نفوذ[1] از جمله این محافظت های امنیتی هستند. اعتماد به سیستم فایروال به تنهایی برای جلوگیری از حملات شبکه کافی نیست. این امر به این دلیل است که یک فایروال نمی تواند از شبکه در برابر نفوذ از درگاه های باز مورد نیاز برای خدمات شبکه دفاع کند. از این رو ، یک سیستم تشخیص نفوذ معمولاً برای تکمیل فایروال نصب می شود. سیستم تشخیص نفوذ یکی از ابزارهایی است که سعی در محافظت از سیستم ها در برابر یک متجاوز دارد. سیستم تشخیص نفوذ اطلاعات را از یک شبکه یا سیستم کامپوتر جمع آوری می کند و اطلاعات مربوط به علائم نقض سیستم را تجزیه و تحلیل می کند(Sangkatsanee et al., 2011). چندیدن الگوریتم یادگیری ماشین مانند NaïveBayes ,k نزدیک ترین همسایه و درخت تصمیم و شبکه های عصبی در زمینه تشخیص نفوذ شبکه های کامپیوتری به کار گرفته شده اند (Ahmim et al., 2019; Liu & Lang, 2019; Mukherjee & Sharma, 2012; Serpen & Aghaei, 2018; Xu et al., 2018). مهم‌ترین چالشی که در این روش ها وجود داشته اول اینکه روش های طبقه بندی به خوبی آموزش ندیده اند و اینکه روش انتخاب مقادیر پارامتر های طبقه بندی کننده متناسب نبوده است. از مزیت های روش ماشین بردار پشتیبان[2] این است که حاشیه جداسازی برای دسته‌های مختلف کاملاً واضح است (Kotpalliwar & Wajgi, 2015) از مزیت های دیگر استفاده از ماشین بردار پشتیبان این هست که وقتی فاصله واضحی از جدایی بین طبقات وجود داشته باشد ، بسیار خوب کار می کند و ماشین بردار پشتیبان در مجموعه داده هایی با ابعاد بالا موثرتر است و ماشین بردار پشتیبان حافظه نسبتاً کارآیی دارد. با این حال، یکی از معایب این طبقه بندی کننده این است که عملکرد آن به انتخاب پارامترهای مناسب بستگی دارد. در حال حاضر الگوریتم های هوش جمعی مختلفی مانند الگوریتم کلونی مورچگان (Mehmod & Rais, 2016) ، بهینه سازی ازدحام ذرات(Kuang et al., 2014) ،الگوریتم کلونی زنبور عسل (Enache & Patriciu, 2014; Wang et al., 2010) ،گرگ خاکستری(Al Shorman et al., 2019) ، الگوریتم خفاش (Enache & Sgarciu, 2015) و... برای بهبود پارامتر های ماشین بردار پشتیبان استفاده میشوند. الگوریتم ماشین بردار پشتیبان با این الگوریتم‌ها ترکیب می‌شود تا مدل‌های تشخیص نفوذ بهبودیافته را بسازند. در بیشتر موارد، آنها برای دو فرآیند بهینه سازی استفاده می شوند: انتخاب ویژگی و انتخاب پارامترهای ماشین بردار پشتیبان. الگوریتم خفاش توسط یانگ در سال ۲۰۱۰ ارائه شد(X. S. Yang & Gandomi, 2012). الگوریتم بهینه سازی خفاش یک الگوریتم بهینه سازی تکاملی است که از خصوصیات ردیابی خفاش ها در جستجوی شکار در طبیعت الهام گرفته شده است. به طور کلی استفاده از الگوریتم خفاش مانند بسیاری از الگوریتم های دیگر خواص سادگی و قابلیت انعطاف را دارد و پیاده سازی ان اسان است و این الگوریتم میتواند برای مسائلی که مقیاسی بالا دارند نیز راه حل پیدا کند. دلایل زیادی مبنی بر تاثیرگزار و کارامدی الگوریتم بهینه سازی خفاش وجود دارد. سه مزایا درباره این الگوریتم شامل الف) میزان سازی فرکانس ب) بزرگنمایی خودکار ج) کنترل پارامتر میشود. فاز اول مدل پیشنهادی، پیش پردازش داده ها با استفاده از k – نزدیکترین همسایه و نرمال سازی داده ها با Min-Max است. در مرحله دوم، انتخاب ویژگی توسط الگوریتم خفاش انجام میشود. در مرحله سوم، الگوریتم خفاش برای بهبود پارامترهای ماشین بردار پشتیبان (C و γ) استفاده می شود. در مرحله چهارم طبقه بندی داده های آموزشی و آزمایشی توسط ماشین بردار پشتیبان بهبود یافته توسط الگوریتم خفاش انجام شده و دقت محاسبه میشود. بر اساس عنوان هایی که

[1] Intrusion Detection System (IDS)

[2] Support Vector Machine (SVM)

بیان شد ، در فصل دوم مرور ادبیات توضیح داده میشود سپس تحقیقات مشابه که در این زمینه انجام شده ، مورد بررسی قرار میگیرند . در فصل سوم مدل پیشنهادی و جزئیات مدل پیشنهادی ارائه میشود . در فصل چهارم به توضیح نحوه پیاده سازی مدل پیشنهادی بر روی مجموعه داده ها و نتایج نهایی مدل پیشنهادی میپردازیم . در پایان نتیجه گیری از کار های صورت گرفته شده در این پایان نامه نتیجه گیری می شود و چند پیشنهاد جهت تحقیقات بیشتر در ادامه این تحقیق انجام شده ارائه میشود.

۲.پیش زمینه

با شروع اولین سیستم تشخیص نفوذ پیشنهاد شده توسط دنینگ[۳] (Denning, 1987)، سیستم های تشخیص نفوذ برای سال های زیادی مورد بحث و توسعه قرار گرفته اند. سیستم های تشخیص نفوذ یکی از زمینه های پرکاربرد یادگیری ماشین به حساب میاید که مهم ترین چالش در این زمینه انتخاب بهترین الگوریتم و روش یادگیری ماشین برای سیستم های تشخیص نفوذ برای رسیدن به مقادیر بهینه و دقت بالا است . به طور کلی، تمام مدل های پیشنهادی دارای مزایا و معایب خود هستند.در این فصل از تحقیق مروری بر الگوریتم های مختلف برای سیستم های تشخیص نفوذ ارائه میشود.

در یک تحقیق (Rajadurai & Gandhi, 2020)مدلی با الگوریتم های PCA-DL پیشنهاد شده است و این الگوریتم بر روی پایگاه داده Nsl-Kdd پیاده سازی شده است.۸۰ درصد داده های اموزشی و ۲۰ درصد تست میباشد و بر روی ان ها پیش پردازش انجام میشود. این مدل از الگوریتم PCA برای تولید اجزا اصلی و از یادگیری عمیق برای تعیین تعداد لایه های ورودی و پنهان استفاده میکند. نتایج نشان می دهد که روش پیشنهادی در مقایسه با الگوریتم های موجود دقت بالاتری را به دست می آورد.

در یک تحقیق (Tharwat et al., 2017) محققان به کمک الگوریتم بهینه سازی خفاش پارامتر های cو گاما را در ماشین بردار پشتیبان بهبود میبخشند. ba-svm بر روی نه مجموعه داده از مجموعه داده های یادگیری ماشین UCI اعمال و برای تایید با ماشین بردار پشتیبان با دو کرنل مقایسه و با الگوریتم های ga-svm و pso-svm مقایسه میشود که ba-svm مقادیر مناسبتری برای پارامتر ها دارد و از مشکل محلی اپتیما جلوگیری می کند و همچنین خطای طبقه بندی کمتری را نسبت به دو الگوریتم دیگر دارد.

گروهی از محققان در (Shen et al., 2018)محققان از الگوریتم یادگیری ماشین افراطی برای طبقه بندی استفاده کرده و برای بهینه سازی از الگوریتم خفاش استفاده شده و یک تابع تناسب بر اساس دقت و تنوع گروه هایی که در الگوریتم خفاش تعریف شده برای بهبود طبقه بندی استفاده میشود. مدل پیشنهادی روی سه مجموعه داده اعمال و میتواند دقت طبقه بندی را افزایش دهد.

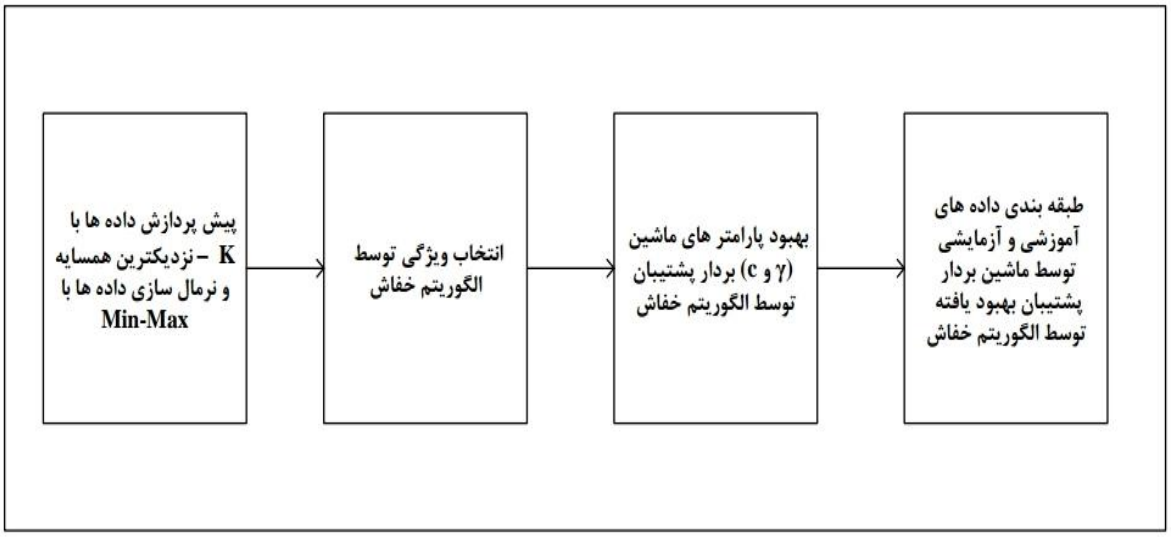
در یک تحقیق دیگر (Nivaashini & Thangaraj, 2019)سیستم تشخیص نفوذ با استفاده از ترکیب خوشه بندی k-means و طبقه بندی ماشین بردار پشتیبان استفاده گردیده است. مدل پیشنهادی از یک مجموعه داده جدید از جریان ترافیک بسته شبکه بی سیم گرفته شده که حالت های مختلف حمله را در خود جای داده است. ترافیک گرفته شده برای تولید یک مجموعه داده ویژه فیلتر و پیش پردازش می شود. نتایج در این تحقیق نشان داده است که false positive به مقدار قابل توجهی کاهش یافته و سرعت تشخیص را افزایش می دهد. شبیه ساز مورد استفاده در این تحقیق (Hall et al., 2009) WEKA می باشد.

[3] Denning

به طور کلی مشاهده می شود که الگوریتم های مختلفی مانند ماشین بردار پشتیبان و بقیه الگوریتم های یادگیری ماشین در مقالات استفاده شده است. با توجه به مطالعاتی که انجام شده می توان بیان نمود که با انجام فاز هایی از قبیل پیش پردازش داده ها و انتخاب ویژگی و طبقه بندی با کمک الگوریتم های یادگیری ماشین میتوان دقت را در سیستم های تشخیص نفوذ افزایش داد. پژوهش های زیادی در این زمینه انجام شده است. اما همچنان کم بودن دقت تشخیص در حملات از مشکلات اصلی ان محسوب میشود.در فصل بعد مدل برای افزایش دقت در تشخیص نفوذ ارائه میشود.

۳.روش پیشنهادی

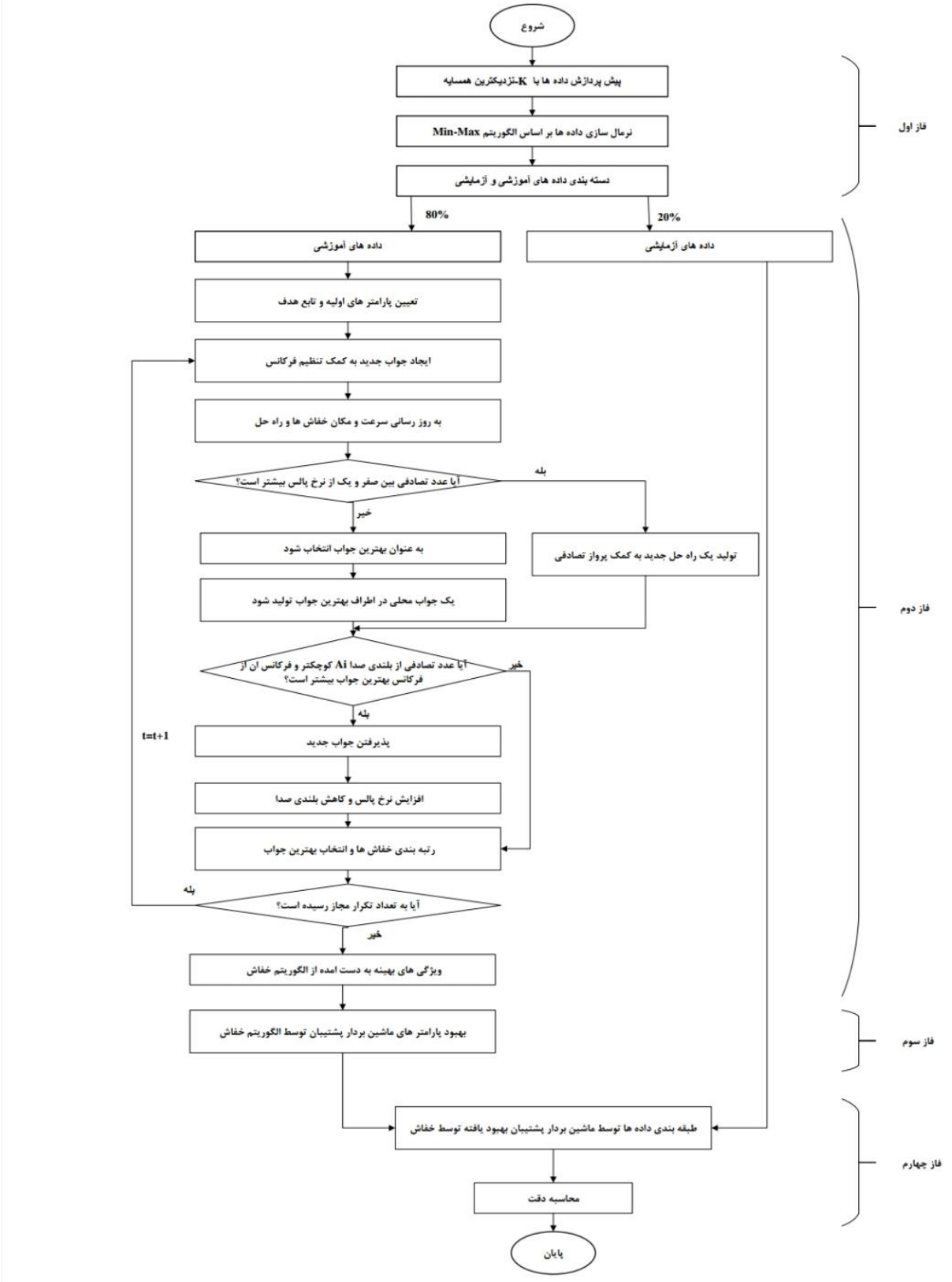
انگیزه ما در تحقیق افزایش دقت در تشخیص نفوذ که یکی از اصلی ترین نکات مورد توجه است میباشد. لذا این تحقیق قصد دارد با ارائه یک مدل برای بهبود پارامتر های ماشین بردار پشتیبان به کمک الگوریتم خفاش و انتخاب ویژگی به کمک الگوریتم خفاش قصد دارد که دقت تشخیص نفوذ را افزایش دهد.مدل پیشنهادی شامل چهار فاز میشود. فاز های کلی مدل پیشنهادی در شکل ۱ ارائه شده است.



شکل۱ - مدل پیشنهادی

در فاز اول پیش پردازش با k – نزدیکترین همسایه (Anbar et al., 2016; Syarif & Gata, 2018) انجام میشود که مقادیر ستون های خالی را با مقادیر نزدیک ترین همسایه جایگذاری میشود. در فاز دوم، انتخاب ویژگی توسط الگوریتم خفاش انجام میشود. ک تکنیک مهم پیش پردازش، انتخاب ویژگی است که بر مدل های یادگیری ماشین تأثیر می گذارد. انتخاب ویژگی با حذف داده های غیر ضروری و غیر مهم از مجموعه داده های اصلی، دقت طبقه بندی را افزایش می دهد. برخی از ویژگی‌ها در یک فضای ویژگی با ابعاد بالا ممکن است اضافی باشند. با حذف این ویژگی‌های اضافی یا نامربوط، عملکرد طبقه‌بندی‌کننده‌ها می‌تواند به طور قابل ملاحظه‌ای بهبود یابد (Eid et al., 2010; Kamel et al., 2019; Zawbaa et al., 2016). در فاز سوم، پارامترهای ماشین بردار پشتیبان توسط الگوریتم خفاش (X. S. Yang, 2013; X. S. Yang & Gandomi, 2012) بهبود یافته است. ماشین بردار پشتیبان یکی از بسیاری از الگوریتم های یادگیری ماشین است. ماشین بردار پشتیبان الگوریتمی برای یافتن بهترین خط یا صفحه بین دو کلاس در فضای ورودی است. در تشخیص الگوهای ظریف در مجموعه

داده های پیچیده موثرتر از سایر الگوریتم های یادگیری ماشین است. برای توسعه سیستم های تشخیص نفوذ قوی، کارآمد و تطبیقی، به دلیل توانایی تعمیم بالا، به حداقل رساندن ریسک ساختاری، انعطاف پذیری در برابر مسائل بیش از حد، و غیره، الگوریتم طبقه بندی ترجیحی بوده است. برای تعریف یک مدل ماشین بردار پشتیبان مناسب، دو پارامتر باید تنظیم شوند: پارامتر حاشیه C و پارامتر آزاد گاما که ضریب کرنل است.هسته تابع پایه شعاعی گاوسی (RBF) یکی از رایج ترین توابع هسته است که شامل پارامتر آزاد γ است. یافتن مناسب‌ترین جفت (C, γ) که برای حل کد مسئله خاص مناسب‌ترین هستند، ضروری است(Gauthama Raman et al., 2019; Tuba & Stanimirovic, 2017). الگوریتم خفاش بر اساس رفتار و پژواک خفاش ها است. مزیت الگوریتم خفاش مدل ساده، همگرایی سریع، موازی بودن بالقوه و ویژگی های توزیع شده آن است (Alsalibi et al., 2020; Yong et al., 2019b). در فاز چهارم، طبقه بندی توسط ماشین بردار پشتیبان بهبود یافته توسط الگوریتم خفاش انجام شده و دقت محاسبه میشود. شکل ۲ جزئیات مدل پیشنهادی را توضیح می دهد.



شکل ۲- جزئیات مدل پیشنهادی

به طور کلی مراحل مدل پیشنهادی شامل چهار فاز میشود که در ادامه جزئیات هر فاز تشریح میشود.

۱-۳- فاز اول : پیش پردازش

این فاز از سه مرحله تشکیل شده است: پیش پردازش ، نرمال سازی و طبقه بندی داده ها . درمرحله اول پیش پردازش داده ها برروی مجموعه داده Nsl-Kdd انجام میشود.دراین تحقیق پیش پردازش داده ها با روش k نزدیک ترین همسایه صورت می گیرد. دراین روش جایگزینی داده ها با مقادیر NAN با نزدیک ترین ستون همسایه صورت می گیرد (Mahboob et al., 2019). پس از پیش پردازش نرمال سازی داده ها مطابق فرمول (۱) بین اعداد ۱و ۱- صورت می گیرد .

$$z = \frac{v - min_A}{max_A - min_A} (new\_max_A - new\_min_A) + new\_min_A \quad (۱)$$

در معادله بالا، v مقدار مشخصه A از محدوده $[min_A,max_A]$ تا محدوده جدید $[new\_min_A,new\_max_A]$ و Z مقدار ویژگی نرمال شده است. مزایای این الگوریتم عبارتند از: تمام روابط بین مقادیر داده را دقیقاً حفظ می کند. هیچ گونه سوگیری احتمالی در داده ها را تعریف نمی کند(Li & Liu, 2011). پس از نرمال سازی داده ها به دو دسته ۸۰ درصد آموزش و ۲۰ درصد تست تقسیم می گردند .

۲-۳- فاز دوم : انتخاب ویژگی های بهینه توسط الگوریتم خفاش

داده ها پس از نرمال سازی و تبدیل به دو دسته ۸۰ و ۲۰ به ماشین بردار پشتیبان برای طبقه بندی ارسال میشوند. اما در این مدل پیشنهادی به دلیل استفاده از انتخاب ویژگی با الگوریتم بهینه سازی خفاش بعد از نرمال سازی داده ها و تبدیل داده ها به دو دسته ۸۰ و ۲۰ ابتدا انتخاب ویژگی های مهم تر توسط الگوریتم خفاش انجام میشود. انتخاب ویژگی یکی از مواردی است که در یادگیری ماشین مورد بحث قرار میگیرد. انتخاب ویژگی در بسیاری از موارد مانند طبقه بندی اهمیت بالایی دارد زیرا در هر مجموعه داده تعداد بالایی ویژگی وجود دارد که تعداد زیادی از ان ها بلا استفاده هستند و یا اطلاعات مناسبی را در اخیار محققان قرار نمیدهد. در صورت حذف نشدن داده ها و ویژگی های اضافی از نظر اطلاعات مشکلی را ایجاد نمیکند اما مشکل به وجود امده برای ان اضافه شدن بار محاسباتی بالا است و هزینه محاسبات را بالا میبرد.زیرا که یکسری ویژگی های غیر مفید را در کنار داده های مفید جای میدهیم و با حذف ویژگی های اضافی میتوان کارایی مدل پیشنهادی را بالا برد پس ویژگی هایی که قدرت تمایز کمتری دارند حذف شده و ویژگی هایی که اطلاعات مفیدی دارند باقی میمانند. از مزایای استفاده از انتخاب ویژگی میتوان به کاهش زمان اموزش و تست اشاره کرد. به طور کلی در انتخاب ویژگی تلاش میشود تا با کاهش حجم داده ها و انتخاب تنها برخی از ویژگی های مجموعه داده اولیه ابعاد ان به طور چشمگیری کاهش پیدا کند و این کار توسط الگوریتم خفاش صورت می گیرد . از مزایای انتخاب ویژگی میتوان به (الف) افزایش درک داده ها، (ب) کاهش اندازه گیری، (ج) نیازهای ذخیره سازی، (د) تسریع فرآیند آموزش (ه) کاهش ابعاد برای بهبود عملکرد پیش بینی اشاره کرد.

۳-۳- فاز سوم: بهبود پارامتر های ماشین بردارپشتیبان توسط خفاش

ماشین های بردار پشتیبان یکی از روش های یادگیری با ناظر است که جزو الگوریتم های یادگیری ماشین محسوب میشوند . ماشن بردار پشتیبان از دقت تعمیم دهی بالا را دارا است. ایده اصلی در ماشین بردار پشتیبان این است که با فرض توانایی جدا شدن داده ها از هم به کمک یک خط ، ابرصفحاتی که قادر به جدا نمودن کلاسها از هم باشند را بدست میاورد. برای ماشین بردار پشتیبان از هسته های متفاوتی مانند RBF و LINEAR میتوان استفاده نمود.که در مدل پیشنهادی از RBF استفاده کردیم. دراین فاز به منظور طبقه بندی داده ها در تشخیص نفوذ به بهبود پارامترcو گاما در هسته RBFتوسط الگوریتم خفاش می پردازیم. فرض میکنیم x نمونه ای از فضای ویژگی ورودی با ابعاد N باشد توابع کرنل یا هسته برای تبدیل

غیرخطی است که داده ها را از یک فضای N بعدی به یک فضای M بعدی نگاشت میکند. کرنل RBF رایج ترین نوع کرنل در الگوریتم ماشین بردار پشتیبان می باشد که دلیل این امر پاسخ های محلی ومحدود آن ها در کل محدوده بردار X مطابق فرمول (2) می باشد :

$$K(x, y) = \exp(-\gamma ||x-y||2) \quad (2)$$

برای $-\gamma>0$ گاهی اوقات از پارامتر $\gamma = \frac{1}{2\sigma^2}$ استفاده می شود که فرمول آن (3) به شرح زیر میباشد:

$$K(x, y) = \exp\left(-\frac{||x-y||^2}{2\sigma^2}\right) \quad (۳)$$

σ یک پارامترآزاد و نشان دهنده پهنای کرنل تابع RBF می باشد. در توابع RBF پارامتری که باید تعیین شود، پارامتر c و γ میباشد . هدف این دو پارامتر این است که مقادیر بهینه دو پارامتر c و γ برای کاربرد موردنظر مشخص شود به گونه ای که ماشین بتواند داده های تست را با دقت مناسبی پیشبینی کند(Ghorbani et al., 2016) . برای طبقه بندی داده ها و افزایش دقت در تشخیص نفوذ از ماشین بردار پشتیبان با پارامتر های بهینه شده C و γ که توسط الگوریتم خفاش به دست امده است ، استفاده میشود . در مدل پیشنهادی به پارامتر های مکان خفاش (x_i)بر حسب دو متغیر C و γ و سرعت (v_i) و فرکانس (f_i) در مکان(x_i)و نرخ پالس (r_i) و بلندی صدا (A_i) که $i=1,2,3,\ldots n$ به در ابتدا صورت تصادفی مقداری برای هر پارامتر در نظر گرفته میشود . هر یک از خفاش ها با توجه به موقعیت مکانی خود به عنوان پارامتر های C و γ ماشین بردار پشتیبان در نظر گرفته میشوند و به عنوان طبقه بندی برروی مجموعه داده اموزشی مورد یادگیری قرار میگیرند. پس از ازمایش برروی مجموعه داده اموزشی اجرا میشود و دقت طبقه بندی به عنوان معیار ارزیابی هر یک از خفاش محاسبه میشود. موقعیت هر خفاش i بر حسب پارامتر های C و γ در زمان t با استفاده از رابطه (4) به تعیین میشود.

$$x_i^t = x_i^{t-1} + v_i^t \quad (4)$$

پس از تعیین موقعیت هر یک از خفاش ها فرکانس صدای تولید شده و سرعت خفاش ها در زمان t با استفاده از روابط (5) و (6) به روز رسانی میشود.

$$f\ i = fmin + (fmax - fmin)\ \beta \quad (5)$$

$$v_i^t = v_i^{t-1} + (x_i^t - x^*) \quad (6)$$

که در رابطه (۵) $\beta \in [0,1]$ و به صورت تصادفی است و x^* موقعیت بهترین خفاش از نظر دقت طبقه بندی ماشین بردار پشتیبان با هسته تابع اساس شعاعی بر روی مجموعه داده ازمایشی با پارامتر های C و γ خفاش مورد نظر میباشد. در اکثر تحقیق های انجام شده در مورد الگوریتم فراابتکاری خفاش برای بهینه سازی الگوریتم ها (Fayaz & Kim, 2018; X. S. Yang & Gandomi, 2012; Yong et al., 2019a) مقدار دو فرکانس $f_{min}=0$, fmax=100است و در مدل پیشنهاد شده نیز از همین مقادیر ها استفاده میکنیم. به منظور پیاده سازی تولید با نرخ پالس (r_i) یک عدد تصادفی بین صفر و یک

تولید میشود . در صورتی که عدد از نرخ پالس (r_i) بزرگتر باشد یک جواب محلی اطراف بهترین جواب با استفاده از رابطه (7) تولید میشود.

$$x_{new} = x_{old} + \epsilon < A^{t+1} > \quad (7)$$

که در رابطه (۷) ϵ یک عدد تصادفی در بازه [-1,1] و $A^t = < A_i^t >$ میانگین بلندی اصوات همه خفاش ها در تکرار t می باشد. همچنین بلندی صوت A_i و نرخ پالس ارسالی r در هر تکرار به کمک فرمول های (8)و (۹)محاسبه می شود. که در آن αو γ ضرایب ثابت معادله اند و برای هر $0 < \alpha < 1$ و$r0>$ وقتی $t \to \infty$ مطابق رابطه (۱۰) داریم:

$$A_i^{t+1} = \alpha A_i^t \quad (۸)$$

$$r_i^{t+1} = r_i^0 \ [1 - \exp(-\gamma t)] \quad (۹)$$

$$A_i^{t+1} \to 0, \ r_i^{t+1} \to r_i^0 \quad (۱۰)$$

بلندی صوت اولیه A_0 میتواند $A_0 \in [1, 2]$ و نرخ پالس ارسالی اولیه نیز $r^0 \in [0, 1]$ میباشد (Gandomi et al., 2013) (Chakri et al., 2017) . زمانی که تعداد تکرار در مراحل اجرایی به اندازه مجاز تعیین شده برسد یا همگرایی انجام شده باشد و از حاصل به دست امده از C و γ بهترین خفاش به عنوان پارامتر های تعیین شده برای ماشین بردار پشتیبان استفاده میشود و دقت طبقه بندی داده ها با استفاده از اجرای ماشین بردار پشتیبان بر روی مجموعه داد ازمایشی محاسبه و مورد تحلیل و بررسی های لازم قرار میگیرد .در غیر این صورت به t یک واحد اضافه شده و فرایند دوباره تکرار میشود.

۴-۳- فازچهارم : طبقه بندی توسط ماشین بردار بهبود یافته توسط الگوریتم خفاش و محاسبه دقت

در این مرحله داده های اموزشی وارد فاز طبقه بندی توسط الگوریتم ماشین بردار پشتیبان بهبود یافته توسط خفاش میشوند. همچنین در این مرحله داده های تست نیز که شامل بیست درصد داده ها میشود . توسط الگوریتم ماشین بردار پشتیبان بهبود یافته توسط الگوریتم خفاش طبقه بندی میشود. در این فصل روش پیشنهادی برای حل مسئله بیان شده ارائه شد.با توجه به این که روش پیشنهادی مبتنی بر الگوریتم خفاش است هر یک از پارامتر ها و نوع عملکرد الگوریتم خفاش و همچنین ماشین بردار پشتیبان مورد بررسی قرار گرفت.

۴.ارزیابی تحقیق

مجموعه داده ای که در این تحقیق مورد بررسی قرار میگیرد مجموعه داده Nsl-Kdd(Tavallaee et al., 2009) است. مجموعه داده Nsl-Kdd (Tavallaee et al., 2009) برای حل برخی از مشکلات مجموعه داده Kddcup99 (Bolón-

Kddcup99 (Bolón-Canedo et al., Canedo et al., 2011; Divekar et al., 2018) پیشنهاد شده است. مجموعه داده Kddcup99 (Bolón-Canedo et al., 2011; Divekar et al., 2018) بیشترین کاربرد را برای تشخیص نفوذ مبتنی بر ناهنجاری دارد اما Tavallaee و همکاران تجزیه و تحلیل آماری در مورد این مجموعه داده انجام دادند و دو مسئله مهم را پیدا کردند که عملکرد سیستم را بسیار تحت تأثیر قرار میداد و نتیجه ارزیابی تشخیص نفوذ در ناهنجاری بسیار ضعیف بود. برای حل این مسائل ، مجموعه داده های Nsl-Kdd (Tavallaee et al., 2009) ، که شامل رکورد های انتخاب شده Kddcup99 است، انتخاب شد(S. Revathi, 2013). مجموعه داده Nsl-Kdd (Tavallaee et al., 2009) شامل ۱۲۵۹۷۳ رکورد و ۴۲ ویژگی یا فیلد است که عبارتند از: ۴۱ ویژگی عادی مربوط به اتصالات شبکه و یک ویژگی کلاس که در آن ۵ کلاس مختلف شامل یک کلاس عادی و ۴ کلاس حمله(Y. Yang et al., 2019) تعریف شده است. کلاسهای حمله عبارتند از: DoS ، U2R، R2L وProbe (Dhanabal & Shantharajah, 2015)روش پیشنهادی در این تحقیق توسط نرم افزار Matlab اجرا شده است.

جهت ارزیابی مدل از ماتریس اغتشاش استفاده شد که یکی از معیار های رایج برای ارزیابی مدل است. عناصر این ماتریس به صورت زیر است :

TN: تعداد دفعاتی که حالت نرمال وجود داشته و مدل پیشنهادی هم ان را به طور درست پیشبینی کرده است.

TP: تعداد دفعاتی که حالت حمله وجود داشته و مدل پیشنهادی هم ان را به طور درست پیشبینی کرده است.

FP: تعداد دفعاتی که حالت نرمال وجود داشته و مدل پیشنهادی ان را به طور اشتباه حمله معرفی کرده است.

FN: تعداد دفعاتی که حالت حمله وجود داشته و مدل پیشنهادی ان را به طور اشتباه نرمال معرفی کرده است.

accuracy یا دقت: به معنای نزدیکی مقادیر اندازه گیری شده به مقدار واقعی است و مدل تا چه اندازه خروجی را درست پیشبینی کرده است. که توسط فرمول (۱۱) قابل محاسبه خواهد بود.

$$accuracy = (TP + TN)/ (TP + TN + FP + FN) \quad (۱۱)$$

در مورد مدل به دست امده با انتخاب ویژگی توسط الگوریتم خفاش و ماشین بردار پشتیبان بهبود یافته توسط الگوریتم خفاش با توجه به مقادیر بالا دقت به دست امده در روش پیشنهادی و در مقایسه با PCA-DL (Rajadurai & Gandhi, 2020) مطابق جدول۱ میباشد:

جدول ۱ - مقایسه دقت روش پیشنهادی با تحقیق پایه

دقت	مجموعه داده	روش پیشنهادی	منبع
۹۲.۴۴	**Nsl-Kdd**	انالیز مولفه اصلی + یادگیری عمیق	**PCA-DL**(Rajadurai & Gandhi, 2020)
۹۷.۸۶	**Nsl-Kdd**	الگوریتم خفاش + ماشین بردار پشتیبان	مدل پیشنهادی

ما مدل خود را با PCA-DL مقایسه کردیم. نتایج جدول ۱ نشان می دهد که مدل پیشنهادی ما میزان دقت را بهبود می بخشد. همچنین سایر معیار های ارزیابی شامل الف) صحت : ۹۷.۹۴ ٪ ب) فراخوانی ۹۵.۴۴٪ ج)معیار f1 :۹۲.۸۵ ٪ د)حساسیت ۶۹.۳۴ همانطور که مشاهده شد مدل پیشنهادی در سایر معیار های ارزیابی نیز به درصد های قابل قبولی رسیده است.اما در تحقیق پایه تنها معیار دقت مورد بررسی قرار گرفته شده است پس سایر معیار های مدل پیشنهادی با تحقیق پایه قابل مقایسه نیست.

۵. نتیجه گیری و پیشنهادات آینده

سیستم تشخیص نفوذ یکی از ابزارهایی است که سعی در محافظت از سیستم ها و تشخیص استفاده های غیر مجاز از سیستم را در برابر تهدید های امنیتی دارد. سیستم های تشخیص نفوذ اطلاعات را از یک شبکه یا سیستم کامپوتر جمع آوری می کند و اطلاعات مربوط به علائم نقض سیستم را تجزیه و تحلیل می کند تا از تکرار در حملات مشابه جلوگیری کند و اطلاعات لازم را درباره حملات انجام شده و نفوذ جمع اوری میشود. هدف این تحقیق ارائه یک مدل پیشنهادی برای بهبود در دقت طبقه بندی در سیستم های تشخیص نفوذ است و در نتیجه سیستم تشخیص نفوذ میتواند تشخیص خرابکاری های در حال وقوع روی شبکه را شناسایی می کند را بهبود ببخشد. به منظور بررسی عملکرد مدل پیشنهادی در سیستم های تشخیص نفوذ مجموعه داده تست بر روی مجموعه داده اموزشی اعمال شد و هر یک از پارامتر های ماتریس اغتشاش محاسبه و سپس معیار های مختلف ارزیابی محاسبه شدند. نتایج معیار های ارزیابی با نتایج معیار های ارزیابی تحقیق پایه (Rajadurai & Gandhi, 2020) مقایسه شد و مدل پیشنهادی نسبت به تحقیق پایه در معیار دقت بهبود بالایی داشته است که از ۹۲.۴۴ درصد به ۹۷.۸۶ درصد رسیده شده است. بهبود دقت به تصمیم گیری بهتر برای تشخیص نفوذ کمک میکند . همچنین نتایج نشان میدهد که مدل پیشنهادی نیز در معیار های دیگر ارزیابی به درصد قابل قبولی رسیده است .

برای تحقیق و پژوهش های اینده موارد زیر پیشنهاد میشود که میتوان الگوریتم های دیگری را در تنظیم پارامترهای ماشین بردار پشتیبان به کار برد و میتوان در پژوهش های اتی از مجموعه داده های دیگر هم استفاده نمود.

منابع:

Ahmim, A., Maglaras, L., Ferrag, M. A., Derdour, M., & Janicke, H. (2019). A novel hierarchical intrusion detection system based on decision tree and rules-based models. *Proceedings - 15th Annual International Conference on Distributed Computing in Sensor Systems, DCOSS 2019*, 228–233.

Al Shorman, A., Faris, H., & Aljarah, I. (2019). Unsupervised intelligent system based on one class support vector machine and Grey Wolf optimization for IoT botnet detection. *Journal of Ambient Intelligence and Humanized Computing 2019 11:7*, *11*(7), 2809–2825.

Almseidin, M., Alzubi, M., Kovacs, S., & Alkasassbeh, M. (n.d.). *Evaluation of Machine Learning Algorithms for Intrusion Detection System*.

Alsalibi, B., Abualigah, L., & Khader, A. T. (2020). A novel bat algorithm with dynamic membrane structure for optimization problems. *Applied Intelligence 2020 51:4*, *51*(4), 1992–2017.

Anbar, M., Abdullah, R., Hasbullah, I. H., Chong, Y. W., & Elejla, O. E. (2016). Comparative performance analysis of classification algorithms for intrusion detection system. *2016 14th Annual Conference on Privacy, Security and Trust, PST 2016*, 282–288.

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2011). Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset. *Expert Systems with Applications*, *38*(5), 5947–5957.

Chakri, A., Khelif, R., Benouaret, M., & Yang, X. S. (2017). New directional bat algorithm for continuous optimization problems. *Expert Systems with Applications*, *69*, 159–175.

Denning, D. E. (1987). An Intrusion-Detection Model. *IEEE Transactions on Software Engineering*, *SE-13*(2), 222–232.

Dhanabal, L., & Shantharajah, S. P. (2015). A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, *4*(6), 446–452.

Divekar, A., Parekh, M., Savla, V., Mishra, R., & Shirole, M. (2018). Benchmarking datasets for Anomaly-based Network Intrusion Detection: KDD CUP 99 alternatives. *Proceedings on 2018 IEEE 3rd International Conference on Computing, Communication and Security, ICCCS 2018*, 1–8.

Drewek-Ossowicka, A., Pietrołaj, M., & Rumiński, J. (2020). A survey of neural networks usage for intrusion detection systems. *Journal of Ambient Intelligence and Humanized Computing 2020 12:1*, *12*(1), 497–514.

Eid, H. F., Darwish, A., Ella Hassanien, A., & Abraham, A. (2010). Principle components analysis and support vector machine based Intrusion Detection System. *Proceedings of the 2010 10th International Conference on Intelligent Systems Design and Applications, ISDA'10*, 363–367.

Enache, A. C., & Patriciu, V. V. (2014). Intrusions detection based on support vector machine optimized with swarm intelligence. *SACI 2014 - 9th IEEE International Symposium on Applied Computational Intelligence and Informatics, Proceedings*, 153–158.

Enache, A. C., & Sgarciu, V. (2015). Anomaly intrusions detection based on support vector machines with an improved bat algorithm. *Proceedings - 2015 20th International Conference on Control Systems and Computer Science, CSCS 2015*, 317–321.

Fayaz, M., & Kim, D. H. (2018). Energy consumption optimization and user comfort management in residential buildings using a bat algorithm and fuzzy logic. *Energies*, *11*(1), 1–22.

Gandomi, A. H., Yang, X.-S., Alavi, A. H., & Talatahari, S. (2013). Bat algorithm for constrained optimization tasks. *Neural Computing and Applications*, *22*(6), 1239–1255.

Gauthama Raman, M. R., Somu, N., Jagarapu, S., Manghnani, T., Selvam, T., Krithivasan, K., & Shankar Sriram, V. S. (2019). An efficient intrusion detection technique based on support vector machine and improved binary gravitational search algorithm. *Artificial Intelligence Review 2019 53:5*, *53*(5), 3255–3286.

Ghorbani, M. A., Zadeh, H. A., Isazadeh, M., & Terzi, O. (2016). A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environmental Earth Sciences 2016 75:6*, *75*(6), 1–14.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18.

Kamel, S. R., YaghoubZadeh, R., & Kheirabadi, M. (2019). Improving the performance of support-vector machine by selecting the best features by Gray Wolf algorithm to increase the accuracy of diagnosis of breast cancer. *Journal of Big Data*, *6*(1), 1–15.

Kotpalliwar, M. V., & Wajgi, R. (2015). Classification of attacks using support vector machine (SVM) on KDDCUP'99 IDS database. *Proceedings - 2015 5th International Conference on Communication Systems and Network Technologies, CSNT 2015*, 987–990.

Kuang, F., Zhang, S., Jin, Z., & Xu, W. (2014). A novel SVM by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection. *Soft Computing 2014 19:5*, *19*(5), 1187–1199.

Li, W., & Liu, Z. (2011). A method of SVM with Normalization in Intrusion Detection. *Procedia Environmental Sciences*, *11*(PART A), 256–262.

Liu, H., & Lang, B. (2019). Machine Learning and Deep Learning Methods for Intrusion Detection

Systems: A Survey. *Applied Sciences 2019, Vol. 9, Page 4396*, *9*(20), 4396.

Mahboob, T., Ijaz, A., Shahzad, A., & Kalsoom, M. (2019). Handling Missing Values in Chronic Kidney Disease Datasets Using KNN, K-Means and K-Medoids Algorithms. *ICOSST 2018 - 2018 International Conference on Open Source Systems and Technologies, Proceedings*, 76–81.

Mehmod, T., & Rais, H. B. M. (2016). Ant Colony Optimization and Feature Selection for Intrusion Detection. *Lecture Notes in Electrical Engineering*, *387*, 305–312.

Mukherjee, S., & Sharma, N. (2012). Intrusion Detection using Naive Bayes Classifier with Feature Reduction. *Procedia Technology*, *4*, 119–128.

Nivaashini, M., & Thangaraj, P. (2019). A framework of novel feature set extraction based intrusion detection system for internet of things using hybrid machine learning algorithms. *2018 International Conference on Computing, Power and Communication Technologies, GUCON 2018*, 44–49.

Rajadurai, H., & Gandhi, U. D. (2020). An empirical model in intrusion detection systems using principal component analysis and deep learning models. *Computational Intelligence*.

S. Revathi, D. a. M. (2013). A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection. *International Jornal of Engineering Research and Technology*, *2*(12), 1848–1853.

Sangkatsanee, P., Wattanapongsakorn, N., & Charnsripinyo, C. (2011). Practical real-time intrusion detection using machine learning approaches. *Computer Communications*, *34*(18), 2227–2235.

Serpen, G., & Aghaei, E. (2018). Host-based misuse intrusion detection using PCA feature extraction and kNN classification algorithms. *Intelligent Data Analysis*, *22*(5), 1101–1114.

Shen, Y., Zheng, K., Wu, C., Zhang, M., Niu, X., & Yang, Y. (2018). An Ensemble Method based on Selection Using Bat Algorithm for Intrusion Detection. *Computer Journal*, *61*(4), 526–538.

Syarif, A. R., & Gata, W. (2018). Intrusion detection system using hybrid binary PSO and K-nearest neighborhood algorithm. *Proceedings of the 11th International Conference on Information and Communication Technology and System, ICTS 2017*, *2018-January*, 181–186.

Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. *IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009*.

Tharwat, A., Hassanien, A. E., & Elnaghi, B. E. (2017). A BA-based algorithm for parameter optimization of Support Vector Machine. *Pattern Recognition Letters*, *93*, 13–22.

Tuba, E., & Stanimirovic, Z. (2017). Elephant herding optimization algorithm for support vector machine parameters tuning. *Proceedings of the 9th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2017*, *2017-January*, 1–4.

Wang, J., Li, T., & Ren, R. (2010). Real time IDSs based on artificial bee colony-support vector machine algorithm. *3rd International Workshop on Advanced Computational Intelligence, IWACI 2010*, 91–96.

Xu, C., Shen, J., Du, X., & Zhang, F. (2018). An Intrusion Detection System Using a Deep Neural Network with Gated Recurrent Units. *IEEE Access*, *6*, 48697–48707.

Yang, X. S. (2013). Bat algorithm: Literature review and applications. *International Journal of Bio-Inspired Computation*, *5*(3), 141–149.

Yang, X. S., & Gandomi, A. H. (2012). Bat algorithm: A novel approach for global engineering optimization. *Engineering Computations (Swansea, Wales)*, *29*(5), 464–483.

Yang, Y., Zheng, K., Wu, C., & Yang, Y. (2019). Improving the Classification Effectiveness of Intrusion Detection by Using Improved Conditional Variational AutoEncoder and Deep Neural Network. *Sensors 2019, Vol. 19, Page 2528*, *19*(11), 2528.

Yao, H., Fu, D., Zhang, P., Li, M., & Liu, Y. (2018). MSML: A Novel Multi-level Semi-supervised Machine Learning Framework for Intrusion Detection System. *IEEE INTERNET OF THINGS JOURNAL*, *XX*.

Yong, J., He, F., Li, H., & Zhou, W. (2019a). A novel bat algorithm based on cross boundary learning and uniform explosion strategy. *Applied Mathematics-A Journal of Chinese Universities*, *34*(4), 480–502.

Yong, J. shi, He, F. zhi, Li, H. ran, & Zhou, W. qing. (2019b). A Novel Bat Algorithm based on Cross Boundary Learning and Uniform Explosion Strategy. *Applied Mathematics-A Journal of Chinese Universities 2021 34:4*, *34*(4), 480–502.

Zawbaa, H. M., Emary, E., & Parv, B. (2016). Feature selection based on antlion optimization algorithm. *Proceedings of 2015 IEEE World Conference on Complex Systems, WCCS 2015*.

Increasing Accuracy of Intrusion Detection System using Support Vector Machine and Bat Algorithm

Mahshid Salehi[1]*, Seyed Reza Kamel [2], Hassan Shakeri[3]
1-Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran. salehy.md@gmail.com
2-Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran. rezakamel@ieee.org
3-Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran. shakeri@mshdiau.ac.ir

Intrusion Detection System (IDS) is a device or software to monitor networks or systems in order to search for malicious activity or policy violations. The most critical challenge in IDS is distinguishing between normal and malicious traffic. The accuracy of IDS has been recently improved through various Learning models. However, the accuracy of the IDS systems still remained a challenge, as attacker and malicious nodes change their behaviors frequently. This research proposes a model to increase the accuracy of the IDS using Support Vector Machine (SVM) and Bat Algorithm (BA). SVM is one of the machine learning algorithms that recently applied by researches for various classification and regression problems and it shows an outstanding performance. Anyhow, the performance of SVM strongly depends on its parameters. In this research, we use BA to optimize the parameters of SVM to increase the accuracy of the IDS. BA has a distinct advantage over other metaheuristic algorithms, BA has the capability of automatically zooming into a region where promising solutions have been found. We evaluate the propose model using Nsl-Kdd dataset. The comparison between one of the recent machine learning algorithms and the present study indicates that the propose model has higher accuracy and better performance than the previous models.