طراحی مدل اجرای مرکز عملیات امنیت (SOC) در صنعت بانکداری

تاریخ دریافت: ۱۴۰۱/۱۱/۳۰
تاریخ پذیرش: ۱۴۰۲/۰۴/۱۴

سید زین العابدین حسینی[۱]، منصور اسماعیل‌پور*[۲]، علیرضا اسلامبولچی[۳]، محمدرضا ربیعی مندجین[۴]، علیرضا امیرکبیری[۴]

۱- دانشجوی دکترای مدیریت فناوری اطلاعات، واحد همدان، دانشگاه آزاد اسلامی، همدان، ایران

۲- دانشیار گروه مهندسی کامپیوتر، واحد همدان، دانشگاه آزاد اسلامی، همدان، ایران (ma_esmaeilpour@yahoo.com)

۳- استادیار، گروه مدیریت، واحد همدان، دانشگاه آزاد اسلامی، همدان، ایران

۴- استادیار گروه مدیریت، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران

چکیده

یکی از مهم‌ترین چالش‌های امنیتی در مراکز عملیات امنیت با الکترونیک، ناتوانی ذاتی اینترنت در مقابله با حملات است. این حملات به‌راحتی اجراشده و به‌صورت محلی یا از راه دور قابل‌کنترل می‌باشند. اکثر این حملات در رسیدن به اهداف اصلی حمله، موفق بوده و مهاجم را به خواسته‌های خود می‌رساند. علت این امر در این است که مکانیسم‌های زیادی برای راه‌اندازی حملات بر اساس مشخصات سرور قربانی وجود دارد، همین امر خود موجب می‌شود که نتوان یک راه‌حل دفاعی جامع در برابر حملات ارائه نمود. راهکارهای متعددی برای شناسایی و مقابله با حملات مزبور ارائه‌شده است که در این مقاله راهکار ترکیب الگوریتم انتخاب ویژگی ژنتیک و روش‌های یادگیری ماشین ازجمله الگوریتم درخت تصمیم، شبکه عصبی عمیق و KNN به‌صورت تلفیقی ارائه شده است. برای اعتبار سنجی راهکار ارائه‌شده، نتایج حاصل با سایر روش‌ها ازجمله روش‌های یادگیری ماشین و ترکیبی با سایر روش‌های بهینه‌سازی مورد مقایسه و ارزیابی شده است. در این پژوهش از ۱۰٪ مجموعه داده KDD Cup 99 برای شبیه‌سازی استفاده‌شده است که ابتدا در مرحله پیش‌پردازش داده‌ها، مقادیر کلیه مشخصه‌ها به اعداد تبدیل و همچنین مقادیر مشخصه خروجی به دو مقدار صفر و یک تغییر داده‌شده است. نتایج حاصل از پژوهش نشان از دقت بالای راهکار ارائه‌شده برای تشخیص نفوذگران نسبت به سایر روش‌های اخیر در حدود ۵٪ است.

واژه‌های کلیدی: **مرکز عملیات امنیت، بانکداری الکترونیک، فرآیند کاوی، یادگیری ماشین**

A New Model of the Security Operations Center (SOC) in the Banking Industry

Seyed Zin El Abidine Hosseini[1], Mansour Esmaeilpour,[*2], Alireza Slambolchi[3], Mohammad Reza Rabiee Mondjin[4], Alireza Amirkabiri[4]

1. PhD. Student, Dept. of Information Technology Management, Islamic Azad University, Hamedan Branch, Hamedan, Iran
2. Assoc. Prof., Dept. of Computer Engineering, Islamic Azad University, Hamedan Branch, Hamedan, Iran (ma_esmaeilpour@yahoo.com)
3. Assist. Prof., Dept. of Management, Islamic Azad University, Hamedan Branch, Hamedan, Iran
4. Assist. Prof., Dept. of Management, Islamic Azad University, Central Tehran Branch, Tehran, Iran

Abstract

One of the most important security challenges in e-banking security centers is the inability of the internet to deal with attacks. These attacks are easily implemented and can be controlled locally or remotely. Most of these attacks are successful in reaching the main targets of the attack and bring the attacker to their desires. The reason for this is that there are many mechanisms for launching attacks based on the characteristics of the victim's server, which makes it impossible to provide a comprehensive defense solution against the attacks. Several strategies have been proposed to identify and deal with these attacks. In this paper, a combination of algorithm for selecting genetic features and machine learning methods, including decision tree algorithm, deep neural network and KNN, are presented. Provide guidelines for validation, the results obtained with other techniques such as machine learning techniques and combined with other optimization methods are compared and evaluated. In this research, 10% of KDD Cup 99 dataset for simulation has been used. First, in the preprocessing of data, the values of all attributes are converted to numbers, and the output characteristic values are changed to two values of zero and one. The results of the research indicate that the accuracy of the proposed strategy for detecting intruders compared to other recent methods is about 5%.

Keywords: Security Operations Center, Electronic Banking, Mining Process, Machine Learning.

۱- مقدمه

مفهوم ایمنی و امنیت از همان آغاز زندگی بشر وجود داشته و بشر همیشه برای بقا و ادامه زندگی سعی نموده که برای حفاظت از خود و دارایی‌هایش، آگاهی و دانش خود را نسبت به محیط و خطرات اطراف خود افزایش دهد. امنیت اطلاعات شامل حفاظت اطلاعات و سیستم‌های اطلاعاتی از فعالیت‌های غیرمجاز مانند دسترسی، استفاده، افشاء، خواندن، نسخه‌برداری یا ضبط، خراب کردن، تغییر، دستکاری و غیره. امنیت اطلاعات به محرمانگی، یکپارچگی و در دسترس بودن داده‌ها مربوط بوده و به فرم اطلاعات اعم از الکترونیک بودن یا چاپی بودن آن مرتبط نیست. دولت‌ها، مراکز نظامی، شرکت‌ها، مؤسسات مالی، بیمارستان‌ها و مشاغل تخصصی مقدار زیادی اطلاعات محرمانه در مورد کارکنان، مشتریان، محصولات، تحقیقات و وضعیت مالی خود گردآوری می‌کنند که بسیاری از این اطلاعات در حال حاضر بر روی کامپیوترها و سرورها جمع‌آوری، پردازش و ذخیره‌شده و به سایر سرورها در شبکه منتقل می‌شود. اگر اطلاعات محرمانه در مورد مشتریان و یا امور مالی یا محصول جدید موسسه به دست رقیب یا سارق اطلاعات بیفتد، این درز اطلاعات ممکن است به خسارات مالی به کسب‌وکار، پیگرد قانونی و یا حتی ورشکستگی یک کسب‌وکار منجر شود. حفاظت از اطلاعات محرمانه یک نیاز تجاری و در بسیاری از موارد نیز یک ضرورت اخلاقی و قانونی است.

ترتیب فناوری‌ها، پروتکل‌ها و روش‌هایی که «امنیت سایبری» نامیده می‌شوند، برای محافظت در برابر حملات، آسیب‌ها، بدافزارها، ویروس‌ها، هک کردن، سرقت داده‌ها و دسترسی غیرمجاز به شبکه‌ها، دستگاه‌ها، برنامه‌ها و داده‌ها است. حفاظت از دارایی‌های کاربر هدف اصلی امنیت سایبری در بانکداری است. همان‌طور که افراد بیشتری بدون پول نقد می‌شوند، اقدامات یا تراکنش‌های اضافی آنلاین می‌شوند. افراد با استفاده از روش‌های پرداخت دیجیتالی مانند کارت‌های نقدی و اعتباری، تراکنش‌ها را انجام می‌دهند که باید توسط امنیت سایبری محافظت شوند. جرائم سایبری در چند سال گذشته به‌طور مکرر افزایش‌یافته است تا جایی که تصور می‌شود یکی از مهم‌ترین خطرات برای بخش مالی است. هکرها فناوری و تخصص خود را بهبود بخشیده‌اند و این امر باعث می‌شود هر بخش بانکی نتواند به‌طور مداوم حمله را خنثی کند. تهدیدات امنیت سایبری به‌طور مداوم در حال تغییر است و بخش بانکی باید برای محافظت از خود اقدام کند. هکرها با توسعه ابزارها و استراتژی‌هایی برای به خطر انداختن امنیت، زمانی که دفاع‌های جدید حملات اخیر را تهدید می‌کنند، سازگار می‌شوند. سیستم امنیت سایبری مالی تنها به‌اندازه ضعیف‌ترین حلقه آن قوی است. داشتن مجموعه‌ای از ابزارها و رویکردهای امنیت سایبری برای محافظت از داده‌ها و سیستم‌های شما بسیار مهم است.

امروزه با توجه به افزایش قابل‌توجه ارزش و اهمیت دارایی‌های موجود در مراکز داده و زیرساخت‌های هر کسب‌وکار و پیرو آن افزایش انواع مختلفی از تهدیدات امنیتی، استفاده از راهکارهای مدیریت و رصد امنیتی به ضرورتی اجتناب‌ناپذیر برای افزایش امنیت و پایداری شبکه و سامانه‌های اطلاعاتی و ارتباطی تبدیل‌شده است. زیرساخت‌های اطلاعاتی سازمان‌های دولتی، شرکت‌های خصوصی و همچنین مراکز عمومی، داده‌های بسیار ارزشمندی را میزبانی می‌کنند که چه‌بسا خرابی هرچند کوتاه و توقف ارائه خدمات توسط آنها، باعث ایجاد خسارات مالی و اعتباری قابل‌توجه خواهد شد. از سوی دیگر انواع مختلفی از حملات و تهدیدات از حملات ویروس‌ها و کرم‌های

اینترنتی گرفته تا حملات ممانعت از سرویس و نفوذ در شبکه همواره زیرساخت‌های اطلاعاتی را تهدید می‌کند. این تهدیدها می‌تواند لطمات و خسارات جبران‌ناپذیری را به داده‌ها و خدمات تحمیل کند [۱].

با رشد روزافزون تهدیدهای امنیتی، پرداختن به مقوله امنیت بیشتر نمایان می‌شود به همین دلیل طی سال‌های اخیر راهکارهای امنیتی داخلی و خارجی متعددی موردتوجه سازندگان، تولیدکنندگان، مؤسسات مالی، کاربران شبکه، مدیران سازمان‌ها و مراکز داده قرارگرفته است. تولید و استفاده تجهیزات امنیتی به‌روز با قابلیت‌های پیشرفته‌تر و تجمیع برخی محصولات این حوزه به‌منظور مدیریت مؤثر تهدیدات و آسیب‌پذیری‌ها ازجمله این راهکارهاست [۲].

در طی چند سال اخیر با توجه به افزایش قابل‌توجه ارزش و اهمیت دارایی‌ها و داده‌های موجود در مراکز داده سیستم‌های اطلاعاتی و به طبع آن افزایش انواع مختلفی از تهدیدات امنیتی، استفاده از راهکارهای مدیریت و رصد امنیتی به ضرورتی اجتناب‌ناپذیر برای افزایش امنیت و پایداری شبکه و سامانه‌های اطلاعاتی و ارتباطی تبدیل‌شده است. راهکاری که اخیراً موردتوجه قرارگرفته است پیاده‌سازی مرکز عملیات امنیت است؛ بنابراین در این مقاله در بخش ۲ کارهای انجام‌شده در گذشته موردبررسی قرارگرفته، در بخش ۳ روش پیشنهادی و جزئیات مربوطه تشریح شده، در نهایت در بخش ۴ نتایج به‌دست‌آمده و در بخش ۵ نیز نتیجه‌گیری نهایی و پیشنهادهای آینده موردبررسی قرار می‌گیرد.

۲- سوابق پیشین

حفظ امنیت اطلاعات در بانکداری موبایل و بستر اینترنت اشیاء و ارائه یک فیلترینگ امن برای جلوگیری از نفوذ و سوء استفاده بر اساس رفتار، سوابق، تراکنش و عملکرد کاربران امروزه بسیار موردتوجه پژوهشگران و نویسندگان مختلف قرارگرفته است. تهدیدات امنیت سایبری به‌طور مداوم در حال تغییر است و بخش بانکی باید برای محافظت از خود اقدام کند. هکرها با توسعه ابزارها و استراتژی‌هایی برای به خطر انداختن امنیت، زمانی که دفاع‌های جدید حملات اخیر را تهدید می‌کنند، سازگار می‌شوند. سیستم امنیت سایبری مالی تنها به‌اندازه ضعیف‌ترین حلقه آن قوی است. داشتن مجموعه‌ای از ابزارها و رویکردهای امنیت سایبری برای محافظت از داده‌ها و سیستم‌های شما بسیار مهم است.

الباهلو و همکارانش در سال ۲۰۱۴، یک چارچوب مبتنی بر GPRS برای امن سازی پروتکل‌های بانکداری موبایل و حفظ امنیت اطلاعات در این دستگاه‌ها ارائه نمودند. آنها در تحقیق خود موفق شدند پروتکل WAP را بهبود بخشیده و موجب امن سازی این پروتکل که یکی از پروتکل‌های پرکاربرد در موبایل است شوند. امن سازی این پروتکل موجب بهبود فرآیند پرداخت و حفظ امنیت اطلاعات در این پژوهش شده است. آنها در تحقیق خود نتایج پیرامون امنیت را موردبررسی قرار دادند. از مهم‌ترین معایب این تحقیق سادگی استراتژی استفاده‌شده و پیچیده کردن مدل روش پیشنهادی است. مهم‌ترین مزایای آن بهبود دقت پروتکل WAP است [۳].

سریرامولو و همکارانش در سال ۲۰۱۵، یک روش جدید بنام STAMBA را به منظور تست امنیت برای برنامه‌های کاربردی بانکداری موبایل آندروید پیشنهاد نمودند و این ابزار را در سطوح مختلف نمایش دادند. این ابزارهای پشتیبانی شده برای پیدا کردن تهدیدها در سطح کد برنامه کاربردی تلفن همراه، ارتباط یا سطح شبکه و در سطح دستگاه مورداستفاده قرار می‌گیرند. آنها در تحقیق خود یک بحث مفصل در مورد آسیب‌پذیری بانکداری موبایل ارائه دادند که به

طراحی برای توسعه برنامه کاربردی بیشتر و یک تست امنیتی خودکار دقیق برای کاربردهای بانکداری موبایل کمک می‌کند. از مهم‌ترین معایب این تحقیق عدم حفظ امنیت اطلاعات پرداخت‌های بانکداری موبایل و داشتن خطای نسبتاً زیاد بر روی موبایل‌ها با سیستم‌عامل آندروید بوده است. ازجمله مهم‌ترین مزایای روش مطرح‌شده در این پژوهش بهبود امنیت بانکداری موبایل با کمک یک استراتژی مبتنی بر اپلیکیشن بوده است [۴].

رفیدها و همکارانش در سال ۲۰۱۶، از یک چارچوب احراز هویت امن در بستر اینترنت اشیاء برای حفظ امنیت اطلاعات در بانکداری موبایل استفاده نمودند. نتایج تحقیقات آنها در این پژوهش منجر به کاهش مصرف انرژی دستگاه‌های موبایل شده و همچنین توانستند به میزان قابل‌توجهی دقت امنیت بانکداری موبایل را نسبت به سایر روش‌ها افزایش دهند. از مهم‌ترین معایب این پژوهش پیچیده بودن مدل و خطای بالای حفظ امنیت اطلاعات است و از مهم‌ترین مزایای این روش نیز سرعت بالای تشخیص حملات، احراز هویت و بهبود مصرف انرژی دستگاه‌های موبایل است [۵].

تامسون و همکارانش در سال ۲۰۱۷، بررسی نمودند که کاربران محاسبات شخصی در برابر تهدیدهای امنیت اطلاعات آسیب‌پذیر هستند، چراکه آنها باید به‌طور مستقل در مورد نحوه محافظت از خود، اغلب با درک کمی از فنّاوری یا مفاهیم آن، تصمیم‌گیری کنند. بااین‌حال، کاربران محاسبات شخصی در مطالعات امنیتی، به‌خصوص برای استفاده از دستگاه‌های تلفن همراه، تحت کنترل هستند. روش تشریح شده در این مقاله این شکاف پژوهشی را با ارزیابی داده‌ها از ۶۲۹ کامپیوتر خانگی و کاربران وسیله تلفن همراه برای بهبود درک رفتار امنیتی در هر دو زمینه موردبررسی قرار می‌دهد. مدل تحقیق تئوری انگیزش محافظت را با توجه به نقش تأثیرات اجتماعی و مالکیت معنوی و ازجمله رفتار واقعی گسترش می‌دهد. این مدل به‌طور جداگانه با کاربران کامپیوتر خانگی و کاربران وسیله تلفن همراه تست شد و داده‌ها نشان می‌دهد که برخی از عوامل تعیین‌کننده رفتار امنیتی بین کامپیوتر خانگی و استفاده از دستگاه تلفن همراه متفاوت است. نتایج نشان می‌دهد که آسیب‌پذیری درک شده، خودکارآمدی، هزینه پاسخ، نرم توصیفی و مالکیت روانی همگی بر مقاصد امنیتی محاسبه شخصی و رفتار کاربران کامپیوتر خانگی و کاربران وسیله تلفن همراه تأثیر می‌گذارند. بااین‌حال، شدت درک شده تنها برای ایفای نقشی در رفتار امنیتی دستگاه تلفن همراه یافت شد و نه اثر واکنش و نه نرم ذهنی بر قصد امنیتی برای هر یک از کاربران تأثیر گذاشت. این یافته‌ها ازنظر مفاهیم کاربردی و پژوهشی و همچنین ایجاد فرصت‌های تحقیقاتی جدید برای امنیت محاسبات شخصی موردبحث قرار می‌گیرند. از مهم‌ترین معایب این تحقیق زمان‌بر بودن فیلترینگ امنیت اطلاعات در بانکداری موبایل و از مهم‌ترین مزایای این پژوهش دقت بالای حفظ امنیت اطلاعات نسبت به تحقیقات پیشین بوده است [۶].

بهاتناگار و همکارانش در سال ۲۰۱۸، بررسی کردند که برنامه‌های بانکداری همراه به دلیل آسیب‌پذیری امنیتی در طراحی برنامه‌های کاربردی و سیستم‌های عامل زمینه‌ای، درخطر حملات سایبری قرار دارند. مکانیسم ارتباط بین فرآیند در آندروید، امکان برقراری ارتباط، اشتراک داده‌ها و قابلیت استفاده مجدد بین آنها را فراهم می‌کند. بااین‌حال، اگر نادرست مورداستفاده قرار گیرد، می‌تواند به یک سطح حمله تبدیل شود که به برنامه‌های مخرب اجازه بهره‌برداری از دستگاه‌ها و به خطر انداختن اطلاعات حساس مالی را می‌دهد. در این تحقیق، آنها بر روی

پرداختن به آسیب‌پذیری هدف با استفاده از یک تکنیک آزمون fuzzing ترکیبی برای تحلیل نیازمندی‌های امنیتی داده‌های کاربردهای مالی آندروید بومی تمرکز نمودند. این سیستم ابتدا به‌طور خودکار یک مدل رفتار برنامه را ایجاد می‌کند و سپس fuzzing ترکیبی را برای مدل تحلیل آسیب‌پذیری نشت داده اعمال می‌کند. نتایج تست به کشف نقاط ورود حملات ناشناخته در برنامه‌های تحت وب کمک می‌کند. ازجمله مهم‌ترین معایب این تحقیق تولید مدل پیچیده به منظور حفظ امنیت بانکداری موبایل بوده است. از مهم‌ترین مزایای این تحقیق بهبود دقت فیلترینگ امنیتی به روش مبتنی بر فازی است [۷].

سن و همکارانش در سال ۲۰۱۸، برای حفظ امنیت اطلاعات در بانکداری موبایل از یک روش تحت عنوان ریسک امنیتی خودکار ارائه نمودند. در این مقاله ابتدا اولین ارزیابی ریسک امنیتی خودکار را انجام داده و تمرکز را روی برنامه‌های بانکی جهانی برای بررسی FinTech است. ابتدا تعداد زیادی از برنامه‌های بانکی مورد تجزیه‌وتحلیل قرارگرفته و مجموعه‌ای جامع از نقاط ضعف امنیتی که در این برنامه‌ها به‌طور گسترده وجود دارد استخراج می‌شود. سپس یک سیستم ارزیابی ریسک امنیتی خودکار سه‌گانه طراحی می‌شود که ترکیبی از پردازش زبان طبیعی و تحلیل استاتیک داده‌ها و جریان‌های کنترل برای شناسایی ضعف‌های امنیتی برنامه‌های بانکی است. آنها در تحقیق خود آزمایش‌های را بر روی ۶۹۳ برنامه بانکی واقعی در بیش از ۸۰ کشور انجام داده و ۲،۱۵۷ ضعف را به استخراج نمودند. تحقیقات آنها نشان داد که نسخه قدیمی برنامه‌های بانکی، آلودگی از کتابخانه‌های شخص ثالث و نیز توابع ضعیف هش، احتمالاً توسط مهاجمان مورد سوءاستفاده قرار می‌گیرند. آنها همچنین نشان دادند که برنامه‌های بانکداران مختلف، انواع مختلف ضعف‌های امنیتی را نشان دارند که عمدتاً به دلیل اقتصاد و مقرراتی که شکل می‌گیرد، است. از مهم‌ترین معایب این تحقیق زمان‌بر بودن فرآیند حفظ امنیت اطلاعات در بانکداری موبایل بوده و از مهم‌ترین مزایای این تحقیق نیز دقت مطلوب تشخیص ناهنجاری‌ها بوده است [۸].

یه و همکارانش در سال ۲۰۱۸، بررسی کردند که چگونگی دستیابی هم‌زمان به استحکام امنیتی و حفظ راحتی استفاده از موبایل در شبکه‌های ارتباطی عمومی ناامن، موضوع مهمی برای تولیدکنندگان دستگاه‌های تلفن همراه هوشمند، شرکت‌های مخابرات و کاربران تلفن همراه است. در این پژوهش، آنها یک طرح تراکنش امن را بنام رمزنگاری مجوز دار برای پرداخت‌های تلفن همراه معرفی نمودند. این طرح پیشنهادی از مزایای استفاده از آندروید و یک سیستم رمزنگاری امضای دیجیتال اصلاح‌شده برای ارائه هم‌زمان امنیت معاملاتی و دستیابی به کارایی پرداخت در عمل استفاده می‌کند. با استفاده از مدل رقیب مشخص‌شده و تجزیه‌وتحلیل امنیتی، ثابت‌شده است که طرح پیشنهادی از طریق مدل اوراکل تصادفی صحیح و ایمن خواهد بود. این سیستم قدرت تبادل قوی و امنیت ارتباطات را برای کاربران تلفن همراه در طول تراکنش‌های پرداخت آنلاین فراهم می‌کند. از طرف دیگر، ارزیابی عملکرد عملی بودن طرح تراکنش پیشنهادی آنها را نشان می‌دهد، چراکه کل هزینه محاسبات برای یک اینترنت مشترک از اشیا مبتنی بر آزمایش قابل‌قبول است [۹].

سوراب و همکارانش در سال ۲۰۱۸، با استفاده از تکنیک‌های یادگیری ماشین اقدام به حفظ امنیت اطلاعات نمودند. آنها در تحقیق خود بررسی کردند که به خاطر تحرک گره‌های شبکه در رایانش ابری سیار، امنیت یک مسئله چالش‌برانگیز بااهمیت بسیار است. هنگامی‌که یک

ابر موبایل شامل شبکه‌های گیرنده ناهمگن، مانند شبکه‌های حسگر بی‌سیم و شبکه‌های بین خودرویی است، مسئله امنیتی بیشتر چالش‌برانگیز می‌شود چون شبکه‌های مشتری اغلب نیازهای امنیتی متفاوتی ازنظر پیچیدگی محاسباتی، مصرف توان و سطوح امنیتی دارند. برای جمع‌آوری اطلاعات و ادغام داده‌های حاصل از شبکه‌های مشتری ناهمگن در سیستم‌های پیچیده این نوع طرح‌های امنیتی جدید نیاز به برنامه‌ریزی دارند. تشخیص نفوذ، یکی از عملکردهای کلیدی امنیتی در یک شبکه متحرک است که شبکه‌های مشتری ناهمگن را درگیر می‌کند. انواع مختلفی از روش‌های تشخیص نفوذ مبتنی بر قاعده را می‌توان در این نوع سیستم‌ها بکار برد. بااین‌حال، طرح‌های کشف نفوذ موجود منجر به پیچیدگی محاسبات بالا یا نیاز به به‌روزرسانی مکرر قانون می‌شوند که به‌طورجدی به کارایی آنها آسیب می‌رساند. در این مقاله، ما یک طرح تشخیص نفوذ مبتنی بر یادگیری ماشین را برای ابرهای متحرک با استفاده از شبکه‌های مشتری ناهمگن پیشنهاد می‌شود. طرح پیشنهادی نیازی به به‌روزرسانی‌های قانون ندارد و پیچیدگی آن را می‌توان مطابق با الزامات شبکه‌های مشتری تنظیم کرد. ازلحاظ فنی، طرح پیشنهادی شامل دو مرحله است: انتخاب ترافیک چندلایه و انتخاب ماشین مجازی مبتنی بر تصمیم (VM). نتایج تجربی ما نشان می‌دهد که طرح پیشنهادی ازنظر تشخیص نفوذ بسیار مؤثر است. از مهم‌ترین مزایای این تحقیق دقت مطلوب و سرعت مناسب برای حفظ امنیت اطلاعات در بانکداری موبایل بوده است و از مهم‌ترین معایب این تحقیق، پیچیدگی پیاده‌سازی است [۱۰].

همان‌طور که از تحلیل‌های صورت گرفته‌شده بر روی تحقیقاتی که تاکنون مطرح‌شده است، مشاهده می‌شود که بسیاری از تحقیقات از عدم داشتن دقت کافی، سرعت فیلترینگ امنیت اطلاعات، عدم داشتن جامعیت لازم و غیره رنج می‌برند؛ بنابراین با وجود چنین گپ (شکاف) و مشکلات عملکردی در مدل‌های مطرح‌شده در زمینه حفظ امنیت اطلاعات در بانکداری موبایل و بستر اینترنت اشیاء، ما در این رساله اقدام به ارائه مدلی خواهیم نمود که تا میزان مطلوبی، دقت حفظ امنیت اطلاعات در بانکداری موبایل و بستر اینترنت اشیاء را افزایش داده و همچنین دارای پتانسیل لازم بر روی تحلیل داده‌های مشکوک باشد. همچنین مدل مطرح‌شده با کمک تکنیک‌های هوش مصنوعی همچون شبکه عصبی عمیق، درخت تصمیم، تکنیک‌های یادگیری تقویتی و غیره، کیفیت امنیت اطلاعات در بانکداری موبایل و بستر اینترنت اشیاء را به‌صورت قابل‌توجهی افزایش داده تا بتواند شکاف‌ها و مشکلات مدل‌ها و روش‌های قبل را بهبود بخشد.

۳- معماری مدل پیشنهادی

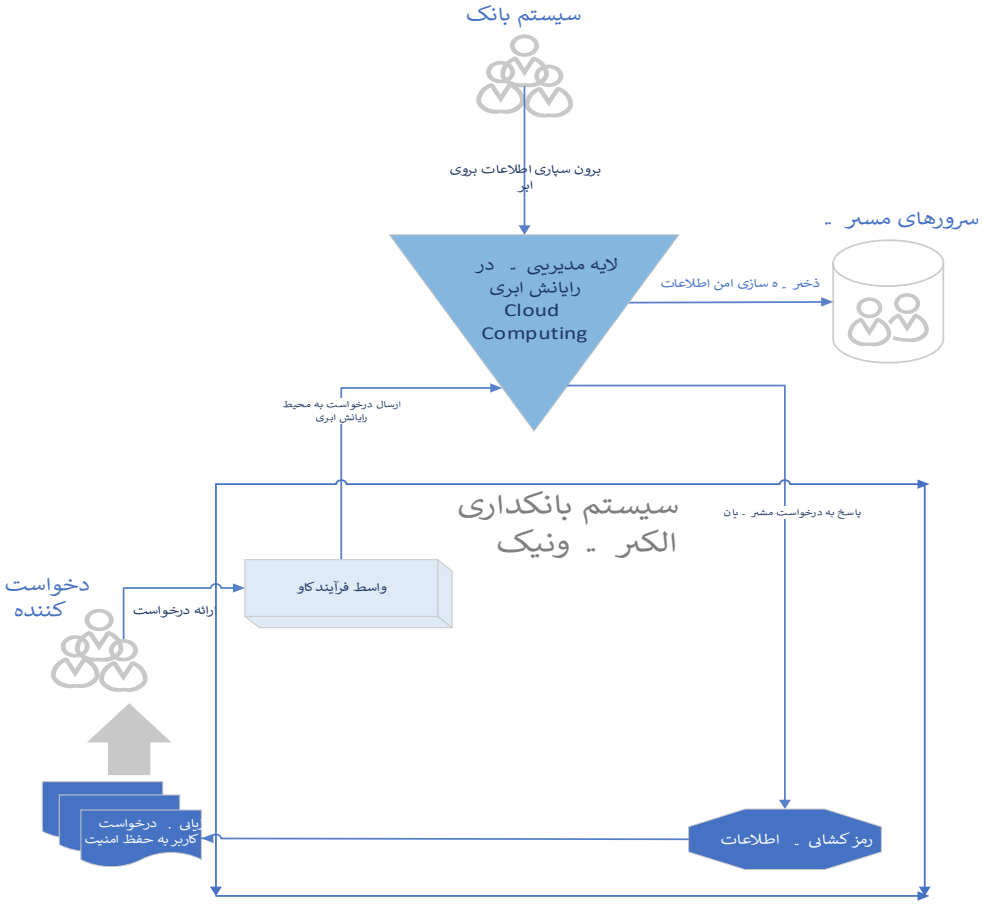
در این بخش معماری کلی مدل اجرای مرکز عملیات امنیت در صنعت بانکداری در لایه مدیریت رایانش ابری نشان داده‌شده است. این معماری بیانگر طرح پیشنهادی این مقاله به‌منظور مرکز عملیات امنیت در صنعت بانکداری است.

همان‌طور که در شکل (۱) دیده می‌شود به‌طورکلی کاربران یا ارسال‌کنندگانی که رویدادهایی را به سمت سرورهای بانک ارسال می‌کنند، افراد یا دستگاه‌هایی هستند که در محیط‌های مختلفی مشغول به فعالیت‌هایی هستند. بانک‌ها اطلاعات مربوط به‌حساب و پرونده‌های مشتریان را از حالت کاغذی به فایل‌های دیجیتالی تبدیل نموده و در رایانش ابری برون‌سپاری می‌کنند. محیط رایانش ابر داده‌ها را رمزنگاری کرده و بروی سرورهای مرکزی و اصلی خود قرار می‌دهد. هر سرور مرکزی نیز شامل یک سری میزبان‌ها است که

امکان ذخیره‌سازی و مدیریت داده‌ها را فراهم می‌سازد. نوع الگوریتم استفاده‌شده به‌منظور رمزنگاری کردن اطلاعات مشتریان در رایانش ابری، الگوریتم استاندارد AES (Algorithm Encryption Standard) است [۱۱]، [۱۲] از مهم‌ترین قابلیت‌های این الگوریتم رمزنگاری، تولید کلیدهای کوتاه، سرعت رمزنگاری بالا و مصرف کم‌حافظه است.

مشتریان سیستم‌های بانکداری الکترونیک که احراز هویت شده و اطلاعات پروفایل و حسابشان در بانک‌های اطلاعاتی سیستم بانک موردنظر (در اینجا، بانک توسعه تعاون) برون‌سپاری شده است، درخواست خود را به لایه مدیریت رایانش ابری در قالب پرسجو ارسال می‌کنند. اطلاعات ارسال‌شده که در قالب بسته‌هایی توسط سیستمی که بین محیط رایانش ابری و سیستم بانکداری الکترونیک قرار دارد، مکانیزه می‌شود. این درخواست‌ها نیز توسط الگوریتم رمزنگاری AES رمز شده و به واسط مربوطه ارسال می‌شود. سیستم واسط بین بانک و محیط رایانش ابری، اطلاعات رمز شده را مورد فرآیند کاوی قرار داده (در ادامه سیستم و مدل فرآیند کاو به‌صورت کامل تشریح می‌شود) و ازلحاظ امنیتی موردبررسی قرار می‌دهد. در نهایت درصورتی‌که درخواست ارسال‌شده دارای رویداد مشکوک باشد، این درخواست به حالت تعلیق در آمده و حساب کاربر ارسال‌کننده مسدود می‌شود. در حالتی که درخواست موردنظر مشکوک نبود به حالت رمزگذاری شده به سمت ابر ارسال می‌شود.

لایه مدیریت رایانش ابر، پرسجوهای دریافتی را به‌صورت امن به رایانش ابری ارسال می‌نماید. پس از بازیابی اطلاعات از منابع داده یا سرورهای



شکل ۱- مدل اولیه ارائه‌شده به‌منظور جانمایی SOC در لایه مدیریت رایانش ابری

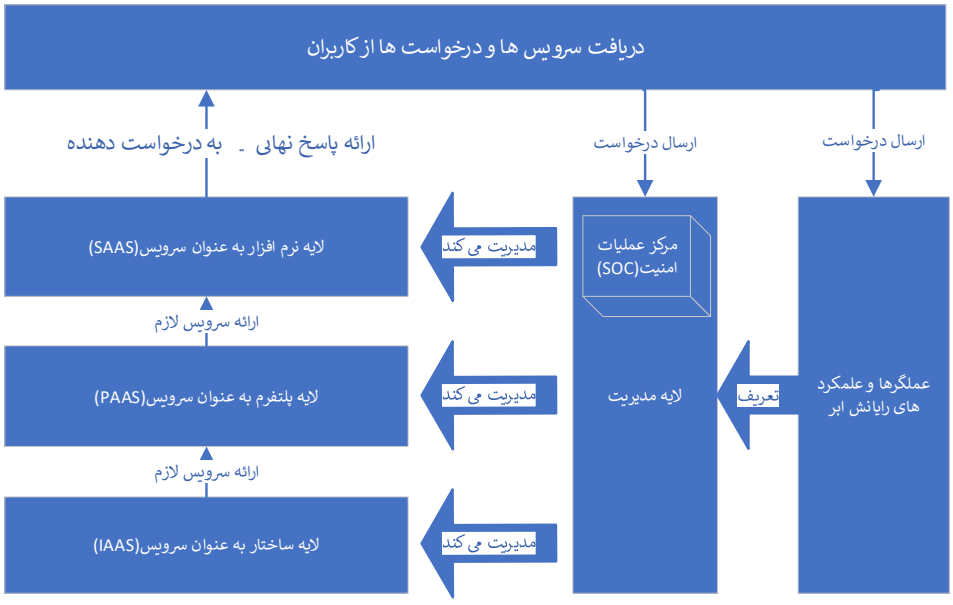
مستر، داده‌های بازیابی شده مجدد به لایه مدیریت در رایانش ابر ارسال‌شده و توسط الگوریتم ADS (Algorithm Decryption Standard) رمزگشایی‌شده [۱۲] و به کاربر نهایی یا سازمان‌های ارائه‌دهنده تحویل داده می‌شود. در بخش بعد قسمت‌های مختلف مدل پیشنهادی تشریح شده و با جزئیات مربوطه موردبررسی قرار می‌گیرد.

با توجه به اهمیت لایه مدیریتی در رایانش ابری، در این قسمت به بررسی این لایه پرداخته و محل جانمایی مرکز عملیات امنیت در صنعت بانکداری مشخص می‌شود. رایانش ابر دارای یک مدل مرجع بوده که در این مدل مرجع لایه‌هایی وجود دارد. یکی از مهم‌ترین و پرکاربردترین لایه‌هایی که به‌منظور جانمایی معماری پیشنهادی استفاده می‌شود، لایه مدیریت در رایانش ابر است. در شکل زیر لایه‌های موجود در معماری مدل مرجع رایانش ابر در مدل پیشنهادی نشان داده‌شده است.

همان‌طور که در شکل (۲) قابل‌مشاهده است، معماری کلی رایانش ابری متشکل از ۵ لایه سرویس مهم است که چارچوب پیشنهادی در لایه مدیریت جانمایی شده است؛ بنابراین به‌طورکلی در شکل قبل می‌توان مدل پیشنهادی را برای اجرای SOC در صنعت بانکداری مورداستفاده قرار داد. در بخش بعد به تشریح مراحل موجود در مدل پیشنهادی پرداخته می‌شود.

۴- تشریح مراحل روش پیشنهادی

مدل مطرح‌شده در بخش قبل دارای یک هسته بوده که در لایه مدیریت رایانش ابری



شکل ۲- جایگاه مرکز عملیات امنیت در مدل مرجع رایانش ابری

پیاده‌سازی شده است. هسته اصلی مدل پیشنهادی را سیستم واسط فرآیند کاو برای حفظ امنیت اطلاعات تشکیل داده است. مراحل مدل پیشنهادی بدین شکل است که در ابتدا سیستم واسط فرایند کاو در نظر گرفته می‌شود که به‌عنوان واسط میان مشتریان و درخواست‌های مطرح‌شده عمل می‌کنند.

۴-۱- سیستم واسط فرآیند کاو

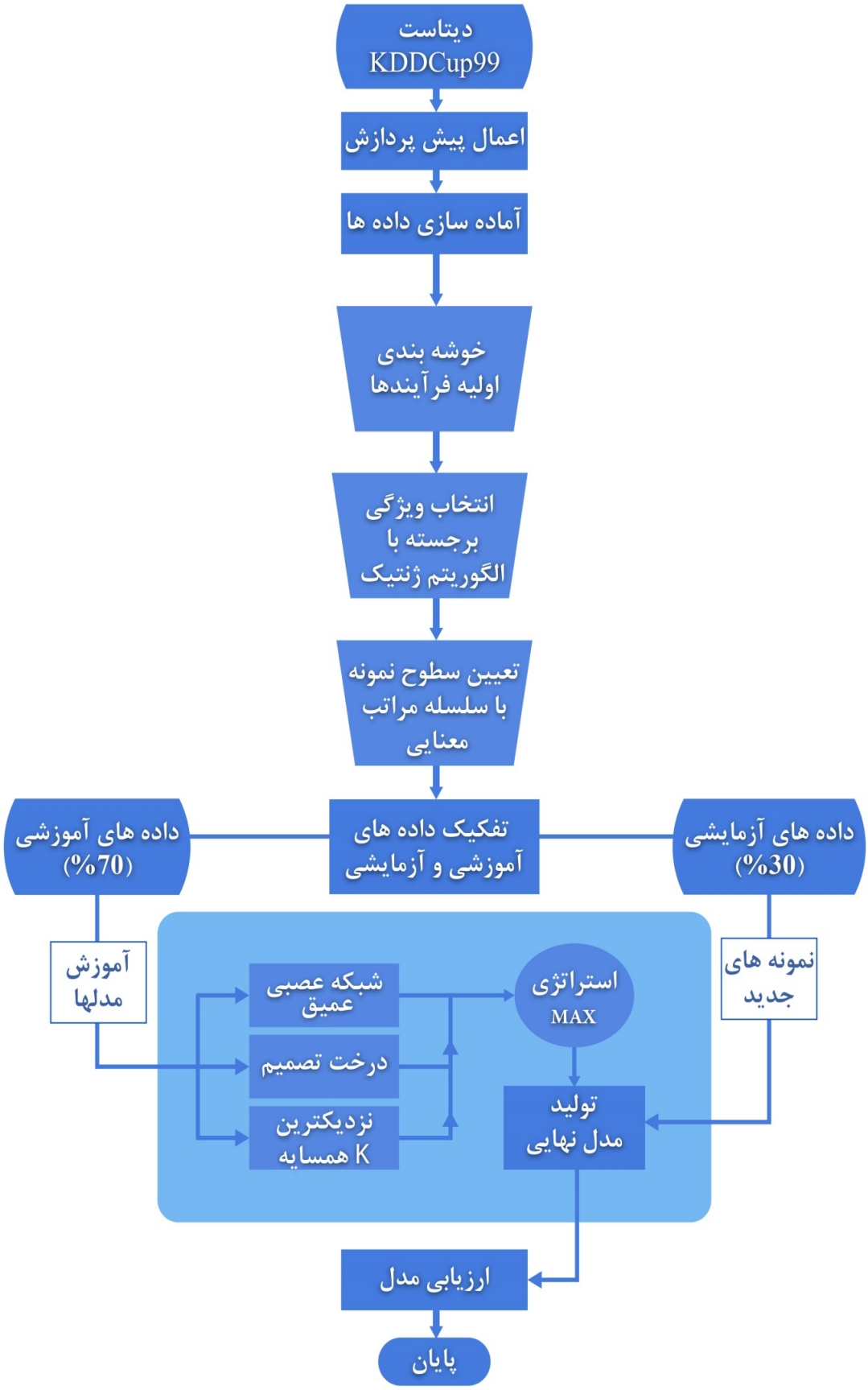
در بین واسطی که مشتریان درخواست‌های خود را ارسال می‌کنند و سرورهای بانک، یک واسط بنام سیستم واسط فرآیند کاو در رایانش ابری

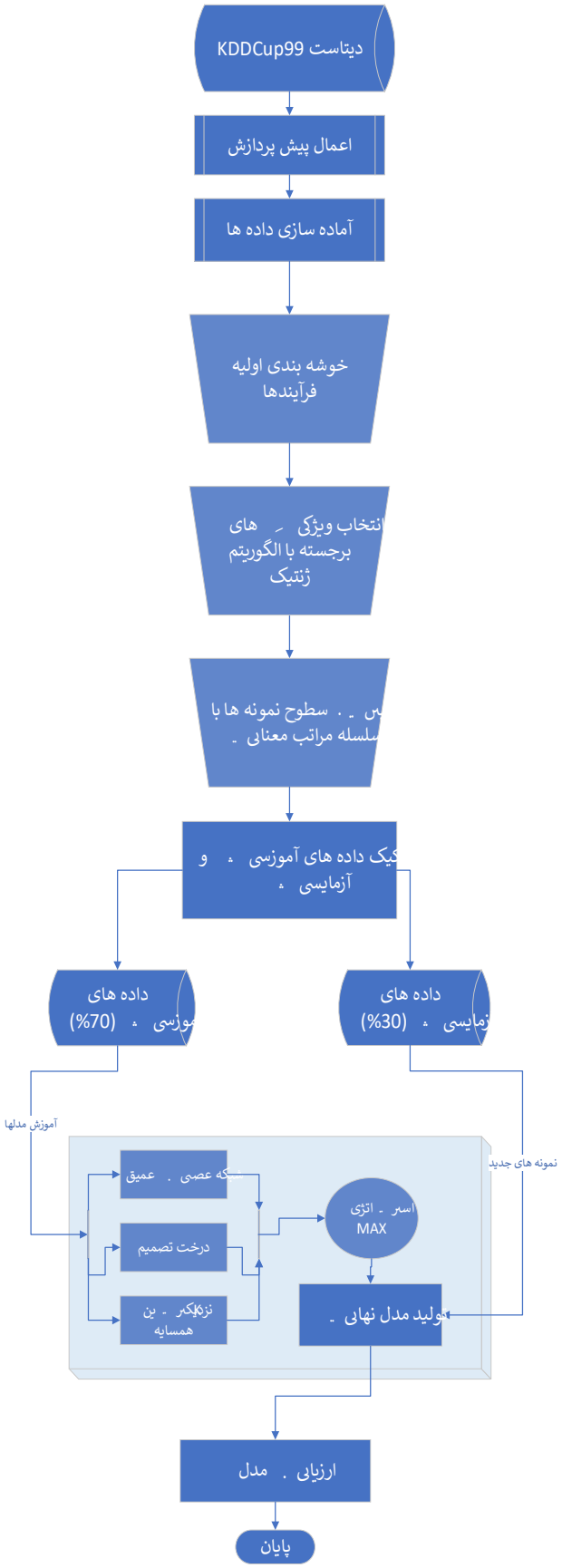
طراحی‌شده است. این واسط وظیفه اصلی اجرایی کردن عملیات امنیت اطلاعات را به عهده دارد. سیستم مطرح‌شده در مدل پیشنهادی، دارای هسته و ساختاری است که در شکل (۳) نشان داده‌شده است.

همان‌طور که از شکل (۳) مشاهده می‌شود، به‌منظور اجرای روش پیشنهادی تا رسیدن به هدف مسئله که اجرای مرکز عملیات امنیت در صنعت بانکداری به‌منظور جلوگیری از نفوذ و حملات به این سیستم‌ها است، ابتدا دیتاست حملات KDDCup 99 به سیستم پیشنهادی وارد می‌شود. ازاین‌رو تمرکز اصلی ما بر روی دیتاست دارپا و حملات دی داس است. سپس این داده‌ها مورد پیش‌پردازش قرارگرفته و داده‌های پرت و بلااستفاده از آن حذف می‌شود. در ادامه نیز داده‌هایی که پس از اعمال پیش‌پردازش به‌صورت منسجم تبدیل شدند، به فرمت قابل‌قبولی برای شبیه‌سازی در لایه مدیریت رایانش ابری تبدیل می‌شود. در این مرحله معمولاً داده‌ها به فرمت اکسل و منسجم تبدیل می‌شود.

در مرحله بعد، داده‌های منسجم شده خوشه‌بندی اولیه می‌شود. الگوریتم خوشه‌بندی استفاده‌شده الگوریتم X-Means است. این الگوریتم نسخه توسعه‌یافته الگوریتم خوشه‌بندی K-Means است. الگوریتم k-Means یک الگوریتم خوشه‌بندی پایه است که بر اساس یک تعداد خوشه بنام k فرآیند خوشه‌بندی نمونه‌ها را انجام می‌دهد. دقت محاسبه برای خوشه‌بندی را می‌توان با مرتب کردن مجدد ردیف‌ها (ستون‌ها) ماتریس سردرگمی انجام داد تا مجموع مقادیر مورب حداکثر باشد. برای هر خوشه، فاصله بین هر یک از اشیاء در خوشه و اشیاء در خوشه‌های دیگر محاسبه می‌شود. از حداقل این فاصله زوجی به‌عنوان جداسازی بین خوشه‌ای (حداقل جداسازی) استفاده می‌شود. فاصله‌ی موردنظر برای هر خوشه و فاصله بین اشیاء در همان خوشه محاسبه می‌شوند. به‌طورکلی اعتبارسنجی الگوریتم خوشه‌بندی در مقایسه با الگوریتم یادگیری ماشینی نظارت‌شده کمی دشوار است زیرا فرآیند خوشه‌بندی حاوی برچسب‌های حقیقت پایه نیست. یکی از مهم‌ترین ایرادات این الگوریتم این است که می‌بایست تعداد خوشه‌ها توسط محقق یا برنامه‌نویس تعیین‌شده و بر اساس این میزان فرآیند خوشه‌بندی انجام شود. تعیین میزان k دارای خطای بالایی بود و اغلب خوشه‌بندی مطلوب و درستی را ارائه نمی‌کرد. به همین دلیل الگوریتم توسعه‌یافته X-Means مطرح شد. این الگوریتم برخلاف الگوریتم K-Means که سرعت بالایی داشته و تعداد k را از ورودی دریافت می‌کرد، دارای سرعت نسبتاً پایینی بوده اما در عوض تعداد k بهینه را خود به دست آورده و تعداد خوشه‌هایی را که دارای کمترین میزان خطا هستند به‌عنوان خوشه‌های پایه منظور می‌نماید. از این تعداد خوشه‌های به‌عنوان ورودی الگوریتم K-Means استفاده کرده و خوشه‌بندی نمونه‌ها را انجام می‌دهد؛ بنابراین، داده‌هایی که مورد پیش‌پردازش قرار گرفتند و منسجم نیز هستند خوشه‌بندی اولیه می‌شود. پس از اینکه داده‌ها خوشه‌بندی اولیه شدند، مقادیر پرت که رفتاری غیره مشابه با سایر نمونه‌ها داشته از سیستم حذف می‌شود. این فرآیند باعث می‌شود تا فرآیند تأمین امنیت سیستم بانکی با دقت بالاتری انجام شود.

در مرحله بعد از روش پیشنهادی، به کمک الگوریتم تکاملی ژنتیک اقدام به انتخاب ویژگی‌های برجسته از مجموعه داده‌های KDDCup99 می‌شود. انتخاب ویژگی‌های برجسته که بیشترین تأثیر در تشخیص حملات


دیتاست
KDDCup99
اعمال پیش پردازش
آماده سازی داده ها
خوشه بندی
اولیه فرآیندها
انتخاب ویژگی
برجسته با
الگوریتم ژنتیک
تعیین سطوح نمونه
با سلسله مراتب
معنایی
داده های آموزشی
(%70)
تفکیک داده های
آموزشی و آزمایشی
داده های آزمایشی
(%30)
آموزش
مدلها
شبکه عصبی
عمیق
درخت تصمیم
نزدیکترین
K همسایه
استراتژی
MAX
تولید
مدل نهایی
نمونه های
جدید
ارزیابی مدل
پایان



شکل ۲- سیستم واسط فرآیند کاو در مدل پیشنهادی

دارد، موجب می‌شود تا اولاً زمان تشخیص حملات نفوذگران کاهش‌یافته و ثانیاً دقت تشخیص نفوذ افزایش یابد؛ زیرا فرآیند آموزش مدل تنها بر اساس ویژگی‌هایی از دیتاست صورت می‌گیرد که دارای بیشترین تأثیر در کشف حملات می‌باشند.

پس از انتخاب ویژگی‌های برجسته با استفاده از الگوریتم ژنتیک، با تعیین سطح معنایی به مقادیر حملات و نمونه‌های معمولی، اقدام به بهبود راه‌حل کشف حمله و اجرای SOC می‌شود. در مرحله بعد، داده با ویژگی‌های برجسته که توسط الگوریتم ژنتیک انتخاب شدند می‌بایست به دو قسمت تقسیم‌بندی شوند که عبارت‌اند از:

- نمونه‌های آموزشی
- نمونه‌های آزمایشی

نمونه‌های آموزشی به‌منظور آموزش روش‌های یادگیری ماشین پیشنهادشده و تولید مدل مربوط به هرکدام استفاده می‌شود. نمونه‌های آموزشی معمولاً ۷۰ الی ۸۰ درصد از نمونه‌ها را تشکیل می‌دهد. نمونه‌های آزمایشی که ۲۰ الی ۳۰ درصد از کل نمونه‌ها را تشکیل می‌دهند به‌منظور ارزیابی و اعتبار سنجی روش ترکیبی استفاده می‌شود. این نمونه‌ها برای تست استفاده‌شده و میزان کارایی و اعتبار سنجی روش پیشنهادی را موردسنجش قرار می‌دهند.

پس از اینکه نمونه‌های آموزشی (۷۰٪) و نمونه‌های آزمایشی (۳۰٪) به روش (Balancing در ادامه تشریح می‌شود) تفکیک شدند، نمونه‌های آموزشی را به الگوریتم‌های یادگیری عمیق، درخت تصمیم C4.5 و الگوریتم K نزدیک‌ترین همسایه به‌عنوان ورودی اعمال می‌کنیم. هرکدام از این الگوریتم‌ها بر اساس نمونه‌های آموزشی مدلی را تولید می‌کنند. این مدل دارای ساختار درختی و ساختار برداری است. سپس نمونه‌های آزمایشی به این مدل‌ها اعمال‌شده و بر اساس نمونه‌های موجود، اعتبارسنجی درخواست‌های مشتریان بانک توسعه تعاون مبتنی بر حداکثر آرا انجام می‌شود. در نهایت نیز روش پیشنهادی مورد ارزیابی قرارگرفته و معیارهایی مثل دقت، صحت، فراخوانی، خطا و غیره مور محاسبه قرار می‌گیرد. در بخش بعد، هسته مدل پیشنهادی مطابق با شکل (۳) تشریح شده و بعدازآن معیارهای ارزیابی نتایج شبیه‌سازی نیز ارائه می‌شود .مطابق با فلوچارت مطرح‌شده در شکل (۷)، اجرای روش پیشنهادی دارای مراحل متعددی است که این مراحل به‌صورت مفصل در ادامه تشریح می‌شود. مهم‌ترین مراحل روش پیشنهادی عبارت‌اند از:

۴-۱-۱- اعمال پیش‌پردازش بر روی داده‌ها

پس از اینکه داده‌ها به سیستم پیشنهادی وارد شدند، داده‌ها موردنظر، پیش‌پردازش می‌شوند و داده‌های پرت و بلااستفاده از آن حذف می‌شود. سپس داده‌هایی که پس از اعمال پیش‌پردازش به‌صورت منسجم تبدیل شدند، به فرمت قابل‌قبولی برای ابزارهای شبیه‌سازی تبدیل می‌شود. در این مرحله معمولاً داده‌ها به فرمت اکسل و منسجم تبدیل می‌شود. روش‌های مختلفی به‌منظور اعمال پیش‌پردازش بر روی داده‌ها مطرح‌شده‌اند که این روش‌ها عبارت‌اند از:

- پاک‌سازی داده
- جمع‌آوری داده
- انتقال داده
- کاهش داده

با توجه به نیاز مسئله ما در این پژوهش ما تنها از روش پاک‌سازی داده‌ها استفاده کرده‌ایم. استراتژی پیشنهادی به این صورت است که داده‌ها را مورد آنالیز قرار داده و درصورتی‌که سطر یا ستونی دارای مقادیر تهی یا بلااستفاده بودند شناسایی می‌شود. سپس مقادیر قبل و بعد از نمونه‌ای که دارای مقدار تهی یا بلااستفاده است را موردبررسی قرار داده و میانگین آنها را محاسبه

مـی‌کنـیم. در نهایـت میـانگین بـه‌دسـت‌آمـده را جایگزین مقدار تهی خـواهیم نمـود. بـا ایـن کـار نمونـه‌هـای پـرت از بـین رفتـه و داده‌هـای منسجم‌تری را تولید خواهیم نمود. همچنین پس از اینکه نمونه‌های پرت از بین رفتنـد مـی‌بایسـت آماده‌سازی داده‌ها را انجـام دهـیم. بـدین منظـور داده‌های پـیش‌پـردازش را بـه فرمـت قابـل‌قبـول ابزارهای شـبیه‌سـازی تبـدیل مـی‌کنـیم. فرمـت پیش‌فرض داده‌ها به‌صورت اکسل است. پس اینکه به‌صورت سـطحی دیتاسـت را مـورد آنـالیز قـرار دادیم می‌بایست به‌صورت رفتاری و مـورد تحلیـل قرار گیرد.

۴-۱-۲- نرمال‌سازی کلان داده‌ها

در مرحله پیش‌پردازش، بـه‌منظـور کسـب نتـایج بهتر، مقادیر هر ویژگی از دیتاست استفاده‌شده را بـین ۰ تـا ۱ نرمـالیزه نمـوده، سـپس سـطرهای ماتریس کلی داده را به‌صـورت تصـادفی جابـه‌جـا مـی‌نمـاییم تـا ترتیـب داده‌هـا از حالـت اولیـه جمع‌آوری‌شده، خارج شود.

به‌عبارتی‌دیگر کلیه دیتاست در قالب ماتریسی نگاشت شده و تغییر سطرهای مـاتریس، عملیـات نرمال‌سازی صورت می‌گیرد. نرمـالیزه نمـودن بـه دلیـل دسـتیابی بـه دقـت بـالاتر اسـت. بـرای نرمال‌سازی مقادیر هـر مجموعـه داده از رابطـه‌ی زیر استفاده‌شده است.

$$Normalize(x) = \frac{(x - X_{\min})}{(X_{\max} - X_{\min})} \qquad (۱)$$

که Xmax و Xmin مقدار بیشـینه و کمینـه در دامنه ویژگی X ام می‌باشند. پس از نرمال‌سـازی داده‌ها، مقادیر کلیه صفت‌هـا در بـازه [۰،۱] قـرار می‌گیرند.

۴-۱-۳- اعمال الگوریتم خوشه‌بنـدی X-Means بر روی داده‌ها

یکی از مهم‌ترین فرآیندهایی که در ایـن پـژوهش انجام می‌شـود اعمـال الگـوریتم خوشـه‌بنـدی X-Means بر روی داده‌ها و کشف نمونـه‌هـایی اسـت که رفتاری مخالف با اغلـب نمونـه‌هـا دارنـد. ایـن نمونه‌ها همان نمونه‌های پرتی هستند که موجـب پیچیده‌تر کردن مـدل در مرحلـه بعـد مـی‌شـود. همان‌طور که گفته شد، الگوریتم X-Means نسخه توسعه‌یافته الگوریتم خوشه‌بندی K-Means است.

الگوریتم k-Means یک الگوریتم خوشـه‌بنـدی پایه است که بر اساس یک تعـداد خوشـه بنـام k فرآیند خوشه‌بندی نمونه‌ها را انجام می‌دهد. یکی از مهم‌ترین ایرادات این الگـوریتم ایـن اسـت کـه مـی‌بایسـت تعـداد خوشـه‌هـا توسـط محقـق یـا برنامه‌نویس تعیین‌شده و بـر اسـاس ایـن میـزان، فرایند خوشه‌بندی انجام شود.

تعیین میزان k دارای خطای بالایی بود و اغلب خوشه‌بندی مطلوب و درستی را ارائه نمی‌کرد. بـه همین دلیل الگوریتم توسعه‌یافته X-Means مطرح شد. این الگوریتم برخلاف الگوریتم K-Means کـه سـرعت بـالایی داشـته و تعـداد k را از ورودی دریافت می‌کرد، دارای سرعت نسبتاً پایینی بـوده اما در عوض تعداد k بهینه را خود به دسـت آورده و تعداد خوشه‌هایی را کـه دارای کمتـرین میـزان خطا هستند بـه‌عنـوان خوشـه‌هـای پایـه منظـور می‌نماید. از این تعداد خوشه‌های به‌عنـوان ورودی الگوریتم K-Means استفاده کرده و خوشـه‌بنـدی نمونه‌ها را انجام می‌دهد؛ بنابراین، داده‌هـایی کـه مورد پیش‌پـردازش قـرار گرفتنـد و منسـجم نیـز هستند خوشه‌بندی اولیه می‌شود.

پس از اینکه داده‌ها خوشه‌بندی اولیـه شـدند، مقادیر پـرت کـه رفتـاری غیـره مشـابه بـا سـایر نمـونه‌هـا داشـته از سیسـتم حـذف می‌شود. این

فـرآیند باعـث می‌شـود تا فـرآینـد اعتبار سنجی درخواست‌های مشتریان بانـک بـا دقـت بـالاتری انجام شود.

در شکل ۴ ساختار داخلی الگوریتم X-Means قابل‌مشاهده است [۱۳]

همان‌طور که از شکل بالا مشـاهده مـی‌شـود، ابتدا تعداد k اولیه تعیین می‌شود. سپس الگوریتم و منطق خوشه‌بندی K-Means به تعداد همـان k تکرار می‌شود. میـزان خطـا محاسـبه مـی‌شـود و سپس یک واحد به تعداد خوشه‌ها اضافه می‌شـود و مراحل قبل مجدد اجرا مـی‌شـود. ایـن روال تـا زمانی ادامه پیدا خواهد کرد که بهترین میـزان k محاسبه شود؛ بنابراین ما در این پـژوهش از ایـن تکنیـک اسـتفاده مـی‌کنـیم تـا بتـوانیم ضـمن خوشه‌بندی اولیـه فرآینـدهـای مشـکوک و غیـره مشکوک، نمونه‌هایی که رفتـار متفـاوت بـا سـایر نمونه‌ها را دارند را شناسایی نمـوده و از مجموعـه داده‌ها حذف کنـیم. در مرحلـه بعـد بـه تفکیـک نمونه‌های آموزشی و آزمایش خواهیم پرداخت.

۴-۱-۴- انتخـاب ویژگـی بهینـه بـا الگـوریتم ژنتیک

از سـال ۱۹۶۰ تقلیـد از موجـودات زنـده بـرای استفاده در الگوریتم‌هـای قدرتمنـد بـرای مسـائل بهینه‌سازی موردتوجه قرار گرفت که تکنیک‌هـای محاسبه تکاملی نام گرفتند درواقع می‌توان گفـت الگوریتم ژنتیک یک تکنیک برنامـه‌نویسـی اسـت که از تکامل ژنتیکی بـه‌عنـوان یـک الگـوی حـل مسئله استفاده می‌کند. هنگامی‌که لغت تنازع بقـا به کار می‌رود اغلب بار ارزشی منفی آن بـه ذهـن می‌آید. شاید هم‌زمان قانون جنگل به ذهن برسـد و حکم بقای قوی‌تر! البتـه بـرای آن‌کـه خیالتـان راحت شود می‌توانید فکر کنید کـه همیشـه هـم قوی‌ترین‌ها برنده نبـوده‌انـد؛ مـثلاً دایناسـورها بـا وجود جثه عظیم در طـی رونـدی کـاملاً طبیعـی بازی بقا را واگذار کرده‌اند درحالی‌کـه موجـوداتی بسیار ضـعیف‌تـر از آنهـا حیـات خـویش را ادامـه داده‌اند. ظاهراً طبیعت بهترین‌ها را تنها بر اسـاس هیکـل انتخـاب نمی‌کند! درواقـع درست‌تر اسـت

```
Input: S: List of segments in MOD, C initialized k cluster
centroids, ¥ Give Threshhold
Output: C_List: List of Clusters
foreach (s ϵ S) do
 foreach (c ϵ C) do
 TempDist=Direction Evaluation (s, direction, c. direction) +
EuclideanDistance (s, c);
 end
 MinDistance=Min[TempDist]. centroid;
 ClosetCentroid=Min[TempDist]. centroid;
 if (MinDistance ≤ ¥) then
 Cluster=C[ClosetCentroid];
 C_List = Update.Centroid(Cluster, s);
 else
 Cluster_New. centroid=s;
 C. Add (Cluster_New. centroid);
end
return C_List;
```

شکل ۳- الگوریتم خوشه‌بندی توسعه‌یافته X-Means [13].

بگوییم طبیعت مناسبترین‌هـا (fittest) را انتخـاب می‌کند نه قوی‌ترین‌ها [۱۴ و ۱۵].

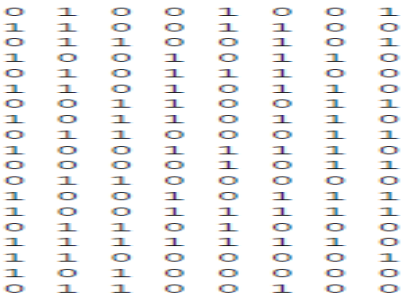
قانون انتخاب طبیعی به این صورت اسـت کـه تنها گونه‌هایی از یک جمعیت ادامه نسل می‌دهند که بهترین خصوصیات را داشته باشـند و آنهـایی که این خصوصیات را نداشته باشند به‌تدریج و در

طی زمان از بین می‌روند. مثلاً فرض کنید گونه خاصی از افراد هوش بسیار بیشتری از بقیه افراد یک جامعه دارند. در شرایط کاملاً طبیعی این افراد پیشرفت بهتری خواهند کرد و این رفاه خود باعث طول عمر بیشتر و باروری بهتر خواهد بود. حال اگر این خصوصیت (هوش) ارثی باشد به طبع در نسل بعدی همان جامعه تعداد افراد باهوش به دلیل زادوولد بیشتر این‌گونه افراد بیشتر خواهد بود. اگر همین روند را ادامه دهید که طی نسل‌های متوالی دائماً جامعه نمونه ما باهوش‌تر می‌شود. به‌این‌ترتیب یک مکانیسم کاملاً ساده طبیعی توانسته است در طی چند نسل عملاً افراد کم‌هوش را از جامعه حذف کنند علاوه بر این‌که میزان هوش متوسط جامعه نیز دائماً در حال افزایش است. بدین ترتیب می‌توان دید که با بهره‌گیری از یک روش بسیار ساده (حذف تدریجی گونه‌های نامناسب و درعین‌حال تکثیر بالاتر گونه‌های بهینه) توانسته است دائماً هر نسل را ازلحاظ خصوصیات مختلف ارتقا بخشد.

روند تکاملی الگوریتم ژنتیک یک شبیه‌سازی ساده و بیولوژیکی است. این تکامل از یک جمعیت تصادفی با توزیع مبتنی بر احتمال شروع می‌شود و معمولاً به‌صورت یکنواخت است که این جمعیت را در مرحله‌هایی به نسل‌ها به‌روز می‌کند. در هر نسل نیز چندین نفر به‌صورت تصادفی از جمعیت فعلی بر اساس یک تابع برازش، ترکیب، جهت و انتخاب به یک جمعیت جدید تبدیل می‌شوند. هر الگوریتم ژنتیک دارای چندین عملگر است. در ادامه ضمن تعریف و کاربرد هر عملگر در الگوریتم ژنتیک، شرح خواهیم داد که هر عملکرد معادل چه بخش‌هایی از قوانین انجمنی و نمونه‌های موجود است.

کروموزوم و ژن

هر کروموزوم در الگوریتم ژنتیک دارای تعداد ژن است. درروش پیشنهادی کروموزوم‌ها دربرگیرنده رکوردهای دیتاست است. در شکل زیر یک مجموعه داده پیش‌فرض نشان داده‌شده است:



شکل ۴- مجموعه داده پیش‌فرض

همان‌طور که مشاهده می‌شود، کروموزوم یکی از رکوردهای موجود در مجموعه داده بالا است. به‌عنوان نمونه: 0 1 0 0 1 0 0 1

و هرکدام از مقادیر ۰ و ۱ یک ژن می‌باشند؛ بنابراین درروش پیشنهادی هر نمونه از دیتاست یا جمعیت اولیه، یک
کروموزوم بوده و هر مقدار از هر نمونه یک ژن است. تعداد کل کروموزوم‌ها معادل با تعداد کل نمونه یا رکوردهای موجود در دیتاست است.

جمعیت اولیه

حجم نمونه یا تعداد کل نمونه‌ها یا داده‌هایی که روش پیشنهادی بر روی آن اجرا می‌شود، جمعیت نام دارد. تعداد جمعیت بر اساس رابطه زیر محاسبه می‌شود [۱۶]:

$$popsize = order[\frac{l}{k} + 2^k] \qquad (۲)$$

که در آن، l تعداد کروموزوم‌ها در دیتاست و k میانگین اندازه شمای دیتاست است. به‌عبارت‌دیگر در پژوهش ما اندازه جمعیت برابر با حجم رکوردهای دیتاست است.

عملگر انتخاب

برای انتخاب بهترین جواب‌ها برای تولید

مجدد نسل (تولید جمعیت جدید) باید از روشی استفاده کرد که حتی‌الامکان بهترین جواب را انتخاب کند. در میان روش‌های مختلف ما چرخ رولت که توسط هالند پیشنهادشده را بررسی می‌کنیم [۱۷]. ایده‌ی اساسی این روش تعیین کردن احتمال بقا برای هر کروموزوم، متناسب با مقدار برازش آن است. چرخ رولت به‌منظور نشان دادن این احتمالات است و فرایند انتخاب مبتنی بر چرخاندن هم‌زمان ارقام چرخ به تعداد اندازه جمعیت است. مقادیر صلاحیت یا V_k مربوط به هر کروموزوم را حساب کنید. به فرض f تابع هدف باشد.

$$eval(V_k) = f(m) = 1.2....pop-size \quad (۳)$$

جمع مقادیر صلاحیت کل کروموزوم‌های موجود را محاسبه کنید

$$F = \sum_{(K=1)}^{(POP-SIZE)} eval(V_k) \quad (۴)$$

احتمال نسبیP_k مربوط به هر کروموزوم را محاسبه کنید

$$p_k = \frac{eval(V_k)}{F} k = 1.2.....pop-size \quad (۵)$$

احتمال تجمعی q_k را برای هر کروموزوم محاسبه کنید

$$q_k = \sum_{(j=1)}^{k} P_j \quad (۶)$$

رویه انتخاب چرخاندن چرخ رولت به تعداد popsize بار آغاز شده است و هر بار یک کروموزوم برای حضور در نسل جدید به‌صورت زیر انتخاب می‌شود:

گام ۱. یک عدد تصادفی مثل r در فاصله [۰،۱] تولید کنید.

گام ۲. اگر r<q_1آنگاه کروموزوم V_1 که اولین کروموزوم است انتخاب می‌شود، در غیر این صورت K امین کروموزوم که در آن۲≥k≤pop- size و q_(k-1)≤r≤q_k است، انتخاب می‌شود.

بنابراین در این پژوهش فرآیند انتخاب بر اساس بهترین میزان برازش در هر بار اجرا صورت می‌پذیرد. رابطه محاسبه میزان برازش در ادامه مطرح می‌شود.

عملگر ترکیب

در این پژوهش از عمل ترکیب یک نقطه‌ای استفاده می‌شود. اگر یک گره معمولی بعد از عمل ترکیب به یک سرگروه تبدیل شود، تمامی گره‌های معمولی دیگر که نزدیک به این نمونه جدید هستند، می‌بایست بررسی شوند و اگر این اتفاق افتاده بود، آنها می‌بایست عضو این دسته جدید شوند.

عملگر جهش

این عملگر در کروموزوم‌های متفاوت تغییرات تصادفی برنامه‌ریزی نشده‌ای ایجاد کرده و ژن‌هایی را که در جمعیت اولیه وجود نداشته‌اند را وارد جمعیت می‌کند. درباره‌ی این عملگر مفهوم مهمی مطرح‌شده که به آن نرخ جهش p_mگویند.

نرخ جهش عبارت است از درصدی از کل تعداد ژن‌های موجود که دچار تغییر می‌شوند. اگر نرخ جهش خیلی کوچک باشد تعداد زیادی از ژن‌هایی که می‌توانستند مفید باشند تست نمی‌شوند؛ اما اگر نرخ جهش خیلی بزرگ باشد نوزادان شباهت‌هایشان را با والدین از دست می‌دهند. این امر باعث از بین رفتن حافظه تاریخی الگوریتم می‌شود. عملگرهای جهش مختلفی وجود دارد، ما فقط نوع یکنواخت آن را

بررسی می‌کنیم. در این عملگر، ژنی از کروموزوم به‌طور تصادفی انتخاب‌شده است و مقدار آن به مقدار تصادفی دیگری تبدیل می‌شود. ابتدا یک عدد تصادفی در بازه [L,۱] که طول کروموزوم موردنظر است و ژن موجود در آن مکان از کروموزوم تغییر می‌کند. فرض کنید کروموزوم والد به‌صورت مقابل باشد

۰۱۰۰۱۰۱۱۰	والد

همان‌طور که مشاهده می‌شود طول کروموزوم ۱۰ است، فرض کنیم عدد تصادفی تولیدشده در بازه [۱۰، ۱] برابر ۵ باشد بنابراین ژن موجود در محل ۵ عوض می‌شود یعنی ۱ تبدیل به ۰ می‌شود

۰۱۰۰۰۰۱۱۰	نوزاد

بنابراین در این پژوهش، در فرآیند جهش نرخ جهش بر اساس تغییر بیت ۰ به ۱ و برعکس است. درنتیجهٔ آن، اگر بیت یک باشد به صفر و اگر صفر باشد به یک تبدیل می‌شود. درواقع یک سرگروه به گره معمولی تبدیل می‌شود و برعکس.

تابع برازش الگوریتم تکاملی ژنتیک برای انتخاب ویژگی

تابع برازش اصلی‌ترین بخش الگوریتم ژنتیک است. در این مقاله از تابع برازش زیر به‌منظور محاسبه میزان برازش هر آیتم استفاده می‌شود.

در رابطه (۷) فرمول برازش نشان داده‌شده است

(۷)

$$fitness_i = w_A \times acc_i + w_F \times \left| 1 - \frac{(\sum_{j=1}^{n_F} f_i)}{n_F} \right|$$

w_A وزن دقت طبقه‌بندی الگوریتم SVM برای انتخاب ویژگی را نشان می‌دهد؛ یعنی به ازای هر ویژگی، یک‌مرتبه الگوریتم SVM اجراشده و میزان دقت محاسبه می‌شود. ⟦acc⟧_i دقت الگوریتم SVM با هسته RBF، w_F وزن برای تعداد ویژگی‌های انتخاب‌شده است، f_i مقدار ماسک ویژگی را نشان می‌دهد که مقدار "۱" نشان می‌دهد که ویژگی j انتخاب‌شده است و مقدار "۰" نشان می‌دهد که ویژگی j انتخاب‌نشده است. n_F تعداد کل ویژگی‌ها را نشان می‌دهد. برای محاسبه ⟦acc⟧_i که دقت طبقه‌بندی SVM را نشان می‌دهد، از رابطه (۸) استفاده می‌شود

$$acc = \frac{cc}{cc + uc} \times 100\% \qquad (۸)$$

cc تعداد نمونه‌هایی که درست طبقه‌بندی‌شده‌اند و uc تعداد نمونه‌هایی که نادرست طبقه‌بندی‌شده‌اند را نشان می‌دهد.

۴-۱-۵- تعیین سطح نمونه‌ها با استفاده از استراتژی سلسله‌مراتب معنایی

روش یا استراتژی سلسله‌مراتب معنایی این امکان را فراهم می‌سازد تا درخواست‌ها، نمونه‌ها و هر نمونه داده‌ای را ازلحاظ محتوا پردازش نموده و به‌صورت معنایی سطح‌بندی یا طبقه‌بندی نماید. تفکیک نمونه‌ها به‌صورت معنایی و مقادیر معنادار این امکان را فراهم می‌سازد تا فرآیند پردازش داده‌ها و تعیین نوع آنها به‌صورت دقیق‌تر صورت پذیرد.

استراتژی سلسله‌مراتب معنایی دارای n سطح معنایی است. هر سطح با توجه به موقعیتی که دارد، مقادیر نمونه‌ها را معنادار می‌کند. سطح معنایی دو، نمونه‌های یک مجموعه داده را بین دو لغت معنادار تقسیم‌بندی می‌کند .

با ذکر یک مثال، سلسله‌مراتب معنایی را شرح

خواهیم داد. فرض شود که مجموعه داده‌های ارسال‌شده مربوط به کاربران سیستم بانکداری است. کاربران دارای ویژگی‌ها و خصیصه‌های سن، جنسیت و شغل هستند. فرض بر این است که این اطلاعات همراه با نمونه‌ها به بستر اینترنت اشیاء ارسال می‌شود. سیستم پیشنهادی یا همان فیلترینگ تقویتی از سطح معنایی ۳ استفاده می‌کند.

سطح معنایی ۳ برای ویژگی سن، سه معنا یا سطح را در نظر گرفته است. اگر سن افراد از بازه [۱۹-۹۵] باشد، استراتژی سلسله‌مراتب معنایی برای افرادی که بازه سن آنها بین ۱۹ تا ۳۵ سال است را به‌عنوان سطح معنایی «جوان» در نظر می‌گیرد. کاربرانی که بازه سن آنها بین ۳۶ تا ۶۰ سال است، توسط استراتژی سلسله‌مراتب معنایی به‌عنوان سطح معنایی «میان‌سال» لحاظ می‌نماید و در نهایت نیز آن دسته از کاربرانی که سن آنها بین ۶۱ تا ۹۵ سال است به‌عنوان افراد «سالخورده» در نظر می‌گیرد.

این استراتژی در مدل پیشنهادی سطح معنایی بهینه‌ای را انتخاب می‌کند. بدین معنا که نمونه‌های جدید با ورود به فیلترینگ تقویتی اقدام به تعیین سطح برای هر نمونه تا n سطح می نماید. سپس به ازای هر سطح، فرآیند طبقه‌بندی نمونه-ها صورت گرفته و هر سطح معنایی که با دقت بالایی اقدام به طبقه‌بندی نمونه‌ها نماید، به‌عنوان سطح معنایی بهینه در نظر گرفته‌شده و در مراحل تست از آن تعداد سطح معنایی استفاده می‌شود.

ابتدا دیتاست پیش‌پردازش شده به الگوریتم سلسله‌مراتب معنایی وارد می‌شود. روش پیشنهادی n سطح را به‌عنوان تعداد سطوح برای سلسله‌مراتب معنایی تعیین می‌نماید. تعداد نمونه های موجود برابر با i و تعداد ویژگی‌های هر نمونه برابر با j در نظر گرفته می‌شود. از ابتدا n=2 فرض شده و سطح معنایی n=2 بر روی ویژگی j از نمونه iاعمال می‌شود. سپس بررسی می‌شود که آیا مراحل اجرایی سلسله‌مراتب معنایی بر روی ویژگی j به اتمام رسیده است یا خیر. اگر به اتمام رسیده باشد سراغ ویژگی j+1 رفته و فرآیند سلسله‌مراتب معنایی را اجرا می‌کند. این عملیات تا زمانی ادامه خواهد یافت که کلیه ویژگی‌ها پیمایش شوند. سپس به نمونه بعدی مراجعه نموده و مراحل قبل مجدد تکرار شده تا کلیه نمونه‌های موجود پردازش شوند .

پس از اینکه سلسله‌مراتب معنایی n=2 بر روی کلیه نمونه‌ها و ویژگی‌ها اعمال شد، یک طبقه‌بندی ساده بر روی کلیه نمونه‌ها با سطح معنایی n=2 اعمال می‌شود. فرآیند طبقه‌بندی با استفاده از الگوریتم درخت تصمیم C4.5 که در ادامه تشریح می‌شود، صورت می‌پذیرد. سپس دقت طبقه‌بندی با سطح معنایی n=2 محاسبه می‌شود. این دقت همراه با سطح معنایی موردنظر در قالب یک ارائه دوبعدی در حافظه ذخیره‌سازی می‌شود. در شکل زیر نمونه‌ای از نتایج به‌دست‌آمده تا سطح معنایی ۵ نشان داده‌شده است.

جدول ۱- نمونه‌ای از نتایج به‌دست‌آمده تا سطح معنایی ۵

سطح معنایی	دقت طبقه‌بندی
n=2	٪۹۰
n=3	٪۸۷
n=4	٪۹۸,۵
n=5	٪۶۹
n=6	٪۸۱,۲

تا این مرحله میزان دقت و سطح معنایی موردنظر در حافظه موجود بوده و یک واحد به سطح معنایی اضافه می‌شود و مجدداً مراحل قبل اجرا می‌شود. این مراحل تا زمانی که به حد آستانه n برسیم تکرار می‌شود. در نهایت بین دقت‌های به‌دست‌آمده یا Max اعمال‌شده و

ماکسیمم دقت محاسبه می‌شود که مربوط به سطح معنایی n=4 است.

لازم به ذکر است که این فرایند در پس‌زمینه لایه Core در شبکه اینترنت اشیاء بر روی نمونه‌های موجود اعمال‌شده تا بتوان تعداد سطح معنایی بهینه را محاسبه نمود. سپس بر اساس سطح معنایی بهینه، در هسته فیلترینگ تقویتی، فرآیند حفظ امنیت اطلاعات ارائه می‌شود. لازم به ذکر است که این فاز از هسته فیلترینگ تقویتی در مدل پیشنهادی می‌تواند قبل از اعمال الگوریتم انتخاب ویژگی ژنتیک باشد. در بخش بعد به تشریح الگوریتم انتخاب ویژگی ژنتیک پرداخته می‌شود.

۴-۱-۶- تفکیک نمونه‌های آموزشی و آزمایشی

نمونه‌برداری از داده‌های موردنظر یکی از مراحل داده‌کاوی است که در این تحقیق موردتوجه قرارگرفته است. روش‌های مختلف نمونه‌برداری وجود دارد که سه مورد از مهم‌ترین آنها عبارت‌اند از [۱۵]

- نمونه‌برداری تصادفی
- نمونه‌برداری طبقه‌بندی‌شده
- نمونه‌برداری متعادل

نمونه‌برداری تصادفی یکی از ساده‌ترین روش‌های نمونه‌برداری است که به‌صورت تصادفی عمل کرده و به میزان موردنظر نمونه‌هایی را از داده‌های اصلی به‌عنوان داده‌های آموزشی و آزمایشی تفکیک می‌نماید. ازجمله معایب این روش این است که ممکن است از یک دسته خاص هیچ نمونه‌ای نمونه‌برداری نشود و در نهایت منجر به کاهش دقت طبقه‌بندی داده‌ها و اعتبار سنجی روش مطرح‌شده می‌شود.

روش نمونه‌برداری طبقه‌بندی‌شده نیز یکی از روش‌های بهبودیافته روش تصادفی است. این روش فرآیند نمونه‌برداری را مبتنی بر احتمال انجام داده و همچنان نمونه‌ها را به‌صورت درصدی انتخاب می‌کند. این روش نمونه‌برداری نیز مشکل داشته و ممکن است نمونه‌هایی را مبتنی بر احتمال انتخاب نکند.

روش نمونه‌برداری متعادل ازجمله روش‌هایی است که نمونه‌های موردنیاز را به‌صورت متعادل از بین دسته‌ها و طبقه‌های موجود انتخاب می‌کند. این روش مشکل روش‌های قبل را نداشته و در نهایت داده‌ها و نمونه‌ها متعادلی را بین نمونه‌های موجود انتخاب می‌کند .

روش استفاده‌شده به‌منظور تفکیک داده‌های آموزشی و داده‌های آزمایشی روش متعادل یا Balancingاست. ازاین‌رو ۷۰٪ از نمونه‌ها به‌عنوان نمونه‌های آموزشی و ۳۰٪ به‌عنوان نمونه‌های آزمایشی تفکیک می‌شود.

۴-۱-۷- اعمال روش پیشنهادی تلفیقی برای اعتبار سنجی

درروش پیشنهادی از محبوب‌ترین روش‌های طبقه‌بندی همچون یادگیری عمیق، درخت تصمیم C4.5 و الگوریتم k نزدیک‌ترین همسایه بهبودیافته، به‌منظور حفظ امنیت اطلاعات درخواست‌ها و فرآیندهایی که توسط مشتریان بانکی ارسال می‌شود، استفاده‌شده است. در نهایت روش‌های فوق باهم ترکیب‌شده و در هر مرحله بهترین پاسخ از بین پاسخ‌های ارائه‌شده انتخاب و به‌عنوان نتیجه نهایی تعیین می‌شود .

با توجه به ماهیت و حساسیت محیط‌های دسته‌بندی فرآیندهای مشتریان اعم از مشکوک و غیره مشکوک، روش پیشنهادی در نظر گرفته‌شده ترکیب شبکه عصبی عمیق، درخت تصمیم C4.5 و روش k نزدیک‌ترین همسایه بهبودیافته است. ترکیب این دو روش در قالب سیستم یادگیری تقویتی ارائه می‌شود.

الگوریتم درخت تصمیم C4.5 و KNN بهبودیافته

الگوریتم درخت تصمیم یکی از مهم‌ترین

الگوریتم الگوریتم‌های یادگیری ماشین است که به‌منظور طبقه‌بندی اطلاعات استفاده می‌شود [۱۸]. از این الگوریتم که در منبع [۱۸] بحث شده است، به‌عنوان یکی از الگوریتم‌ها در سیستم یادگیری تقویتی استفاده می‌شود.

این روش‌ها ازجمله روش‌های محبوب داده‌کاوی هستند که در کنار الگوریتم استاندارد یادگیری عمیق استفاده می‌شود. در الگوریتم KNN بهبودیافته[19] K، [20] به‌عنوان K اصلی منظور می‌شود که دارای بیشترین میزان دقت باشد. در بخش بعد الگوریتم یادگیری عمیق تشریح می‌شود.

الگوریتم شبکه عصبی عمیق

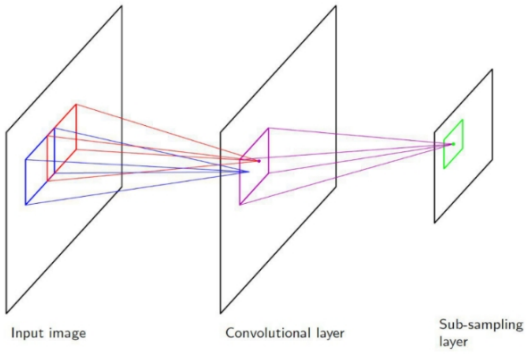
به‌طورکلی در شبکه‌های عصبی یک ورودی دریافت می‌کنند (در قالب یک بردار) و سپس آن را از تعدادی لایه مخفی عبور می‌دهند؛ و نهایتاً یک خروجی که نتیجه پردازش لایه‌های مخفی است در لایه خروجی شبکه ظاهر می‌شود. هر لایه مخفی از تعدادی نورون تشکیل‌شده که این نورون‌ها به تمام نورون‌های لایه قبل از خود متصل می‌شوند.

نورون‌های هر لایه به‌صورت مستقل عمل کرده و هیچ ارتباطی با یکدیگر ندارند. آخرین لایه تماماً متصل به لایه خروجی معروف است و معمولاً نقش نمایش‌دهنده امتیاز هر دسته (class) را ایفا می‌کند. شبکه‌های عصبی معمولی برای داده‌های معمولی به‌خوبی مقیاس‌پذیر نیستند [۲۱].

شبکه‌های عصبی عمیق از این واقعیت که ورودی شامل نمونه‌های معتبری است استفاده کرده و معماری شبکه را به روش معقولی محدود کردند. به‌طور خاص، برخلاف یک شبکه عصبی معمولی، لایه‌های یک شبکه عصبی عمیق (به‌اختصار) شامل نورون‌هایی است که در سه بعد عرض، ارتفاع و عمق قرارگرفته‌اند (مرتب شده‌اند).

(دقت کنید که کلمه عمق در اینجا اشاره به بُعد سوم یک توده فعال‌سازی دارد و به معنای عمق یک شبکه عصبی کامل که به معنای تعداد لایه‌های موجود در آن است نیست .

همان‌طور که جلوتر خواهیم دید، هر نورون در هر لایه بجای اتصال با تمام نورون‌ها در لایه قبل تنها به ناحیه کوچکی از لایه قبل از خود متصل است. علاوه بر آن، لایه خروجی نهایی برای مشتریان بانکی دارای بُعد خواهد بود. چراکه همگام با رسیدن به انتهای معماری شبکه ConvNet ما اندازه نمونه‌ها را کاهش می‌دهیم به‌گونه‌ای که در انتها نمونه‌ها کامل ورودی ما به یک بردار حاوی امتیاز دسته‌ها (کلاس‌ها) کاهش پیدا می‌کند و ما با یک بردار که حاوی امتیاز هر دسته است مواجه خواهیم بود. این امتیازات در امتداد بعد عمق مرتب شده‌اند. نمایشی از این عمل را در زیر می‌توانید مشاهده کنید.



شکل ۵- قسمتی از یک شبکه عصبی عمیق [۲۱].

شکل (۶)، یک قسمتی از یک شبکه عصبی عمیق را نشان می‌دهد که یک لایه از واحدهای عمیقی که با استفاده از یک واحد جایگزینی ادامه دارد را نشان می‌دهد. جفت‌های متعددی از این دنباله لایه‌ها استفاده می‌شود. همان‌طور که در نمونه بالا می‌بینید در هر لایه، یک شبکه عصبی عمیق نورون‌های خود را در ۳ بعد مرتب می‌کند (عرض، ارتفاع و عمق) هر لایه یک شبکه عمیق ورودی را در قالب یک توده سه‌بعدی به یک توده

سه‌بعدی خروجی از مقادیر فعال‌سازی نورون‌ها تبدیل می‌کند. در این مثال لایه ورودی قرمزرنگ حاوی تصویر است (مقادیر پیکسل‌های تصویر) بنابراین عرض و ارتفاع آن ابعاد تصویر خواهند بود و عمق آن‌هم برابر با ۳ خواهد بود (کانال‌های قرمز، سبز و آبی مربوط به تصویر).

یک شبکه عمیق از چندلایه تشکیل می‌شود و هر لایه شیوه کار ساده‌ای دارد که در آن‌یک توده سه‌بعدی ورودی دریافت کرده و آن را با استفاده از توابعی مشتق‌پذیر که ممکن است با پارامتر یا بدون پارامتر باشند به یک توده سه‌بعدی خروجی تبدیل می‌کند.

ازآنجایی‌که مقادیر مربوط به این پارامترهای مراحلی به‌صورت خودکار تنظیم می‌شود، ما از آن به یادگیری یاد می‌کنیم، چراکه شبکه عصبی گام‌به‌گام با یادگیری این پارامترها قادر به انجام وظیفه شناسایی محول شده به آن می‌شود. توده فعال‌سازی یا به یک توده سه‌بعدی حاوی مقادیر عددی گفته می‌شود که به‌عنوان ورودی به تابع فعال‌سازی ارسال می‌شوند، برای همین به آنها توده فعال‌سازی گفته می‌شود.

مقادیر موجود در این توده‌ها ممکن است مقادیر متناظر به نمونه‌های خام داده‌ها باشند (توده فعال‌سازی ورودی) و یا نتیجه پردازش‌های انجام‌شده تا لایه خاصی در شبکه باشند به‌عنوان‌مثال توده فعال‌سازی در لایه دوم یعنی مقادیر عددی در لایه دوم که نتیجه عملیات لایه‌های قبل تا لایه فعلی است (ضرب وزن‌ها در خروجی حاصل از لایه قبل و....)

در اینجا مقادیر فعال‌سازی چیزی جز مقادیر مربوط به پیکسل‌های خام تصاویر ورودی نیستند. همان‌طور که در بالا اشاره کردیم، هر لایه شبکه عمیق یک توده فعال‌سازی را از طریق یک تابع مشتق‌پذیر به توده فعال‌سازی دیگر تبدیل می‌کند؛ بنابراین یکی از روش‌هایی که به‌منظور حفظ امنیت اطلاعات در صنعت بانکداری الکترونیک و به‌خصوص بانک توسعه تعاون استفاده می‌شود الگوریتم شبکه عصبی عمیق است.

۴-۱-۸- اعمال روش یادگیری تقویتی

بر اساس روش ارائه‌شده که ترکیبی از روش‌های عصبی عمیق با تعداد ۲ لایه پنهان، درخت تصمیم C4.5 و الگوریتم طبقه‌بندی KNN است، در این قسمت به تشریح نحوه ترکیب این روش‌ها پرداخته می‌شود.

روال کار به این صورت است که:

۱- داده‌های آموزشی به مدل شبکه عصبی عمیق با تعداد ۲ لایه پنهان، درخت تصمیم C4.5 و الگوریتم طبقه‌بندی KNN وارد می‌شود. نوع الگوریتم شبکه عصبی عمیق استفاده‌شده الگوریتم استاندار شبکه عصبی عمیق بوده است.

۲- هر سه روش بیان‌شده، مدل‌های خود را آموزش داده و آماده دریافت داده‌های آزمایشی به‌منظور شناسایی مشتریان معتبر و نامعتبر هستند.

۳- پس از اینکه مدل‌های مربوطه آموزش داده شدند، داده‌های آزمایشی که ۳۰ درصد از کل داده‌ها هستند برای ارزیابی به مدل‌ها وارد می‌شوند. هرکدام از مدل‌ها خروجی و میزان پیش‌بینی و شناسایی خود را به‌عنوان خروجی برمی‌گردانند.

۴- خروجی این روش‌ها به ورودی هسته سیستم بوستینگ متصل شده و با توجه به پارامتر موجود در سیستم بوستینگ که min، max، Avg است بهترین پیش‌بینی و طبقه‌بندی انتخاب‌شده و به‌عنوان خروجی استفاده می‌شود.

۵- با ارائه روش پیشنهادی فوق در هر بار اجرا و پیش‌بینی مورد جدید بهترین پاسخ به خروجی ارسال‌شده و در نهایت بهینه‌ترین پاسخ به دقت بالا و کمترین خطا را شاهد خواهیم بود. در شکل زیر یک نمای کلی از سیستم یادگیری

تقویتی ارائه شده است.

بنابراین با توجه به دیدگاه ارائه‌شده در این بخش روش پیشنهادی را در بخش بعد شبیه‌سازی نموده و نتایج به‌دست‌آمده را با سایر روش‌ها ارزیابی خواهیم نمود .

۵- نتایج تجربی

۵-۱- مشخصات سیستم و شبیه‌سازی

شبیه‌سازی انجام‌شده در یک سیستم با مشخصات زیر آزمایش‌شده است

جدول ۲- فاکتورها و مشخصات سیستم شبیه‌سازی

فاکتور	مشخصه
اندازه دیسک	۸ گیگ
ظرفیت حافظه RAM	۵۱۲ مگابایت
تعداد پردازنده	۱ و۲ و۴ و۸

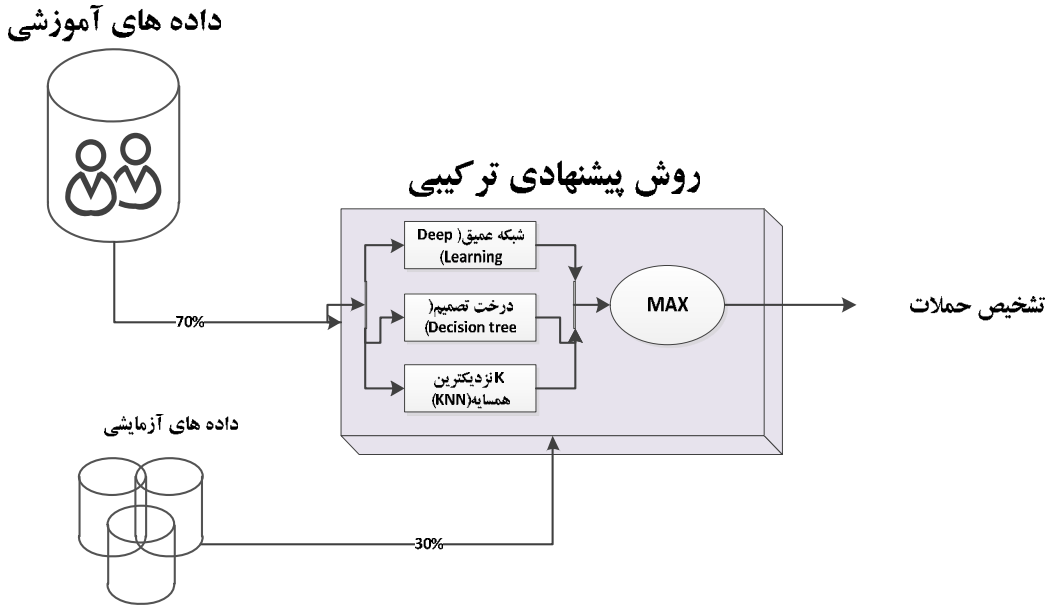
۵-۲- معیارهای ارزیابی نتایج

به‌منظور ارزیابی روش‌های طبقه‌بندی استفاده‌شده برای دسته‌بندی مشتریان معیارهایی وجود دارد که میزان صحت و درستی، خطا و میزان دقت روش‌ها را محاسبه می‌نماید. ازجمله مهم‌ترین این معیارها عبارت‌اند از:

- دقت
- صحت
- فراخوانی
- خطا

در رابطه‌های (۹) روش‌های محاسبه میزان دقت، صحت، فراخوانی و خطای دسته‌بندی نشان داده‌شده است

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (۹)$$



شکل ۶- فلوچارت روش تقویتی پیشنهادی

در فرمول بالا TP (True Positive) بیانگر فرآیندهایی است که نوع دسته‌شان درست بوده و درست نیز تشخیص داده‌شده است TN (True Negative)بیانگر تعداد فرآیندهایی است که نادرست بوده و نادرست تشخیص داده‌شده است .

FP (False Positive)نیز بیانگر تعداد فرآیندهایی است که نادرست بوده است و درست تشخیص داده‌شده است. در نهایت نیز FN (False Negative) نمونه‌هایی را نشان می‌دهد که نادرست بوده و دقیقاً نادرست طبقه‌بندی‌شده است.

با ذکر یک نمونه فرمول فوق واضح‌تر بیان می‌شود. فرض کنید ۱۰ نمونه وجود دارد که مشکوک و غیره مشکوک بودن یک نوع تراکنش بانکی را مشخص می‌کنند. حال فرض کنید یک نمونه فرآیند انجام می‌شود که مشکوک است؛ زمانی که مثلاً الگوریتم درخت تصمیم C4.5 بروی این نمونه اجراشده و می‌خواهد نوع فرآیند را طبقه‌بندی کند، نمونه موردنظر را که مشکوک بوده است را مشکوک طبقه‌بندی نموده است؛ اما شبکه عصبی عمیق مثلاً این فرآیند را نامشکوک طبقه‌بندی کرده است؛ بنابراین یک FN توسط الگوریتم درخت تصمیم C4.5 انجام می‌شود و یک FP نیز توسط شبکه عصبی عمیق نیز اتفاق افتاده است.

FN ارائه‌شده یعنی فرآیند مشکوک (نادرست) بوده و مشکوک نیز تشخیص داده‌شده است و در نهایت FP یعنی تراکنش مشکوک بوده اما نامشکوک تشخیص داده‌شده است. اگر همین عملیات به‌صورت عکس برای تراکنش نامشکوک ارائه شود، TP، TNنیز اتفاق می‌افتد.

TP، FN پاسخ‌های صحیح بوده و FP، TN پاسخ‌های نادرست هستند.

رابطه (۱۰) مربوط به ارزیابی صحت و فراخوانی به ترتیب به شرح ذیل است

$$precision = \frac{TP}{TP + FP} \qquad (۱۰)$$

در فرمول (۱۰) TP بیانگر نمونه‌هایی است که مقادیر نوع دسته‌شان درست بوده و درست نیز تشخیص داده‌شده است. FP نیز بیانگر تعداد نمونه‌هایی است که نادرست بوده است و درست تشخیص داده‌شده است

$$Recall = \frac{TP}{TP + FN} \qquad (۱۱)$$

در نهایت نیز FN نمونه‌هایی را نشان می‌دهد که نادرست بوده و دقیقاً نادرست طبقه‌بندی‌شده است. در نهایت میزان خطا نیز با فرمول (۱۲) محاسبه می‌شود

(۱۲)

$$Error\ rate = 100 - (\frac{TP + TN}{TP + TN + FP + FN})$$

۵-۳- تحلیل نتایج روش پیشنهادی

در این قسمت به مقایسه تشخیص نفوذگران با اعمال الگوریتم بهینه‌سازی ژنتیک و بدون اعمال این الگوریتم و الگوریتم‌های یادگیری ماشین با یکدیگر پرداخته می‌شود .

در شکل زیر مقایسه دقت روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذگران بدون اعمال الگوریتم ژنتیک نشان داده‌شده است

همان‌طور که از شکل (۸) مشاهده می‌شود، میزان دقت روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذ گران بدون اعمال الگوریتم ژنتیک برابر با ۹۹,۸۴٪، روش یادگیری عمیق برابر با ۹۹,۲۵٪، روش نزدیک‌ترین همسایه برابر با ۹۹,۸۱٪ و روش درخت تصمیم برابر با ۹۹,۸٪ است. با این تفاسیر میزان بهبود دقت روش پیشنهادی در مقایسه با روش‌های یادگیری عمیق، نزدیک‌ترین همسایه و درخت تصمیم به ترتیب برابر با ۰/۵۹٪، ۰/۰۳٪ و ۰/۰۴٪ است.

میزان صحت روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذ گران بدون اعمال الگوریتم ژنتیک برابر با ۹۹,۹٪، روش یادگیری عمیق برابر با ۹۸,۴٪، روش نزدیک‌ترین همسایه برابر با ۹۹,۸۸٪ و روش درخت تصمیم برابر با ۹۹,۸٪ است.

با این تفاسیر میزان بهبود صحت روش پیشنهادی در مقایسه با روش‌های یادگیری عمیق، نزدیک‌ترین همسایه و درخت تصمیم به ترتیب برابر با ۱,۵٪، ۰,۰۲٪ و ۰,۱٪ است. میزان

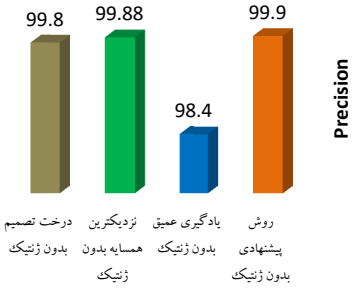
فراخوانی روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذ گران بدون اعمال الگوریتم ژنتیک برابر با ۹۹,۶۳٪، روش یادگیری عمیق برابر با ۹۸,۴۷٪، روش نزدیک‌ترین همسایه برابر با ۹۹,۵۷٪ و روش درخت تصمیم برابر با ۹۹,۶٪ است.

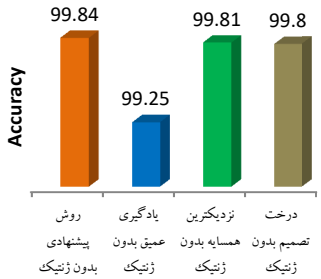
با این تفاسیر میزان بهبود فراخوانی روش پیشنهادی در مقایسه با روش‌های یادگیری عمیق، نزدیک‌ترین همسایه و درخت تصمیم به ترتیب برابر با ۰,۱۶٪، ۰,۰۶٪ و ۰,۰۳٪ است. میزان خطای روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذ گران بدون اعمال الگوریتم ژنتیک برابر با ۰,۱۶٪، روش یادگیری عمیق برابر با ۰,۷۵٪، روش نزدیک‌ترین همسایه برابر با ۰,۱۹٪ و روش درخت تصمیم برابر با ۰,۲٪ است.

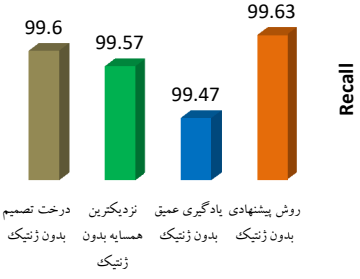
با این تفاسیر میزان بهبود خطای روش پیشنهادی در مقایسه با روش‌های یادگیری عمیق، نزدیک‌ترین همسایه و درخت تصمیم به ترتیب برابر با ۰,۵۹٪، ۰,۰۳٪ و ۰,۰۴٪ است. میزان خطای واقعی روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذ گران بدون اعمال الگوریتم ژنتیک برابر با ۰,۰۰۳٪، روش یادگیری عمیق برابر با ۰,۰۶۴٪، روش نزدیک‌ترین همسایه برابر با ۰,۰۰۲٪ و روش درخت تصمیم برابر با ۰,۰۰۳٪ است.

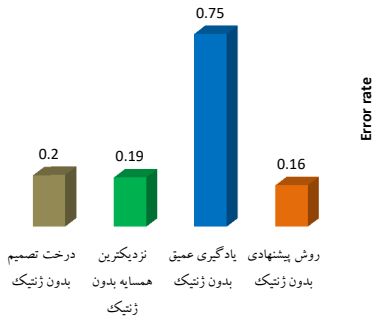
با این تفاسیر میزان بهبود خطای واقعی روش پیشنهادی در مقایسه با روش یادگیری عمیق با۰,۰۶۱٪ است. همچنین میزان مجذور مربعات خطای روش پیشنهادی با سایر روش‌های پایه بدون اعمال الگوریتم ژنتیک برابر با ۰,۰۴۲٪، روش یادگیری عمیق برابر با ۰,۱۵۹٪، روش نزدیک‌ترین همسایه برابر با ۰,۰۳۵٪ و روش درخت تصمیم برابر با ۰,۰۴۵٪ است.

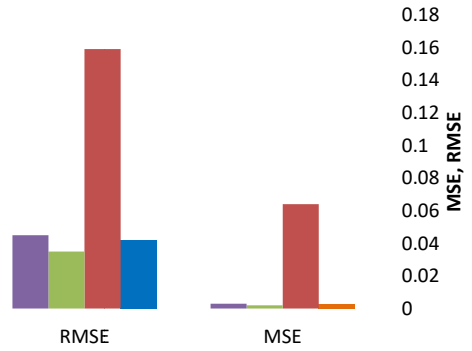
با این تفاسیر میزان بهبود مجذور مربعات خطای روش پیشنهادی در مقایسه با روش‌های یادگیری عمیق و درخت تصمیم به ترتیب برابر با ۰,۱۱۷٪ و ۰,۰۰۳٪ است .









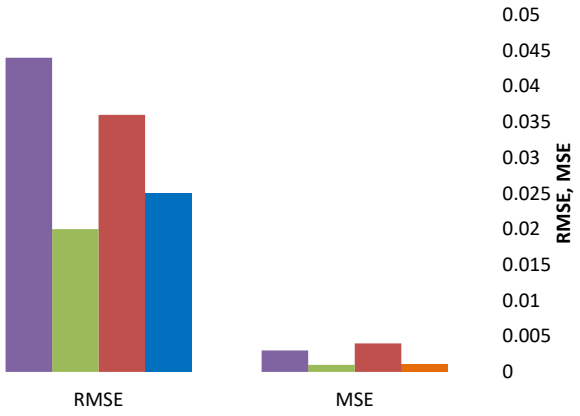


شکل ۷- مقایسه دقت، صحت، فراخوانی، خطا، مجذور مربعات و خطای واقعی روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذگران بدون اعمال الگوریتم ژنتیک

در شکل زیر مقایسه دقت روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذگران با اعمال الگوریتم ژنتیک نشان داده‌شده است.

همان‌طور که از شکل (۹) مشاهده می‌شود، میزان دقت روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذ گران با اعمال الگوریتم ژنتیک برابر با ۹۹,۹۳٪، روش یادگیری عمیق برابر با ۹۹,۹۲٪، روش نزدیک‌ترین همسایه برابر با ۹۹,۹۵٪ و روش درخت تصمیم برابر با ۹۹,۸۱٪ است.

با این تفاسیر میزان بهبود دقت روش پیشنهادی در مقایسه با روش‌های یادگیری عمیق و درخت تصمیم به ترتیب برابر با ۰,۰۱٪ و ۰,۱۲٪ است و روش نزدیک‌ترین همسایه نسبت به روش پیشنهادی ۰,۰۲٪ بهبود داشته است. میزان صحت روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذ گران با اعمال الگوریتم



شکل ۹- مقایسه دقت، صحت، فراخوانی، خطا، مجذور مربعات و خطای واقعی روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذگران با اعمال الگوریتم ژنتیک

ژنتیک برابر با ۹۹,۹۴٪، روش یادگیری عمیق برابر با ۹۹,۹۲٪، روش نزدیک‌ترین همسایه برابر با ۹۹,۹۷٪ و روش درخت تصمیم برابر با ۹۹,۸۶٪ است.

با این تفاسیر میزان بهبود صحت روش پیشنهادی در مقایسه با روش‌های یادگیری عمیق و درخت تصمیم به ترتیب برابر با ۰,۰۲٪ و ۰,۰۸٪ است و روش نزدیک‌ترین همسایه نسبت به روش پیشنهادی ۰,۰۳٪ بهبود داشته است.

میزان فراخوانی روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذ گران با اعمال الگوریتم ژنتیک برابر با ۹۹,۸۵٪، روش یادگیری عمیق برابر با ۹۹,۸۶٪، روش نزدیک‌ترین همسایه برابر با ۹۹,۹٪ و روش درخت تصمیم برابر با ۹۹,۵۹٪ است.

با این تفاسیر میزان بهبود فراخوانی روش پیشنهادی در مقایسه با روش درخت تصمیم برابر با ۰,۲۶٪ است و روش‌های یادگیری عمیق و نزدیک‌ترین همسایه نسبت به روش پیشنهادی به ترتیب برابر با ۰,۰۱٪ و ۰,۰۵٪ بهبود داشته است. میزان خطای روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذگران با اعمال الگوریتم ژنتیک برابر با ۰,۰۷٪، روش یادگیری عمیق برابر با ۰,۰۸٪، روش نزدیک‌ترین همسایه برابر با ۰,۰۵٪ و روش درخت تصمیم برابر با ۰,۱۹٪ است.

با این تفاسیر میزان بهبود خطای روش پیشنهادی در مقایسه با روش‌های یادگیری عمیق

جدول ۳- نتایج کلی الگوریتم‌های شبکه عصبی عمیق، درخت تصمیم، KNN و ترکیب این روش‌ها با اعمال الگوریتم ژنتیک و بدون الگوریتم ژنتیک

	بدون اعمال انتخاب ویژگی						با اعمال انتخاب ویژگی		
	روش پیشنهادی	یادگیری عمیق	نزدیک‌ترین همسایه	درخت تصمیم	روش پیشنهادی	یادگیری عمیق	نزدیک‌ترین همسایه	درخت تصمیم
دقت	99.84	99.25	99.81	99.8	99.93	99.92	99.95	99.81
صحت	99.9	98.4	99.88	99.8	99.94	99.92	99.97	99.86
فراخوانی	99.63	99.47	99.57	99.6	99.85	99.86	99.9	99.59
خطا	0.16	0.75	0.19	0.2	0.07	0.08	0.05	0.19
MSE	0.003	0.064	0.002	0.003	0.001	0.004	0.001	0.003
RMSE	0.042	0.159	0.035	0.045	0.025	0.036	0.02	0.044

و درخت تصمیم به ترتیب برابر با ۰٫۰۱٪ و ۰٫۱۲٪ است. میزان خطای واقعی روش پیشنهادی با سایر روش‌های پایه برای تشخیص نفوذگران با اعمال الگوریتم ژنتیک برابر با ۰٫۰۰۱٪، روش یادگیری عمیق برابر با ۰٫۰۰۴٪، روش نزدیک‌ترین همسایه برابر با ۰٫۰۰۱٪ و روش درخت تصمیم برابر با ۰٫۰۰۳٪ است.

با این تفاسیر میزان بهبود خطای واقعی روش پیشنهادی در مقایسه با روش‌های یادگیری عمیق و درخت تصمیم به ترتیب برابر با ۰٫۰۰۳٪ و ۰٫۰۰۲٪ است. همچنین میزان مجذور مربعات خطای روش پیشنهادی با سایر روش‌های پایه با اعمال الگوریتم ژنتیک برابر با ۰٫۰۲۵٪، روش یادگیری عمیق برابر با ۰٫۰۳۶٪، روش نزدیک‌ترین همسایه برابر با ۰٫۰۲٪ و روش درخت تصمیم برابر با ۰٫۰۴۴٪ است.

با این تفاسیر میزان بهبود مجذور مربعات خطای روش پیشنهادی در مقایسه با روش‌های یادگیری عمیق و درخت تصمیم به ترتیب برابر با ۰٫۰۱۱٪ و ۰٫۰۱۹٪ است.

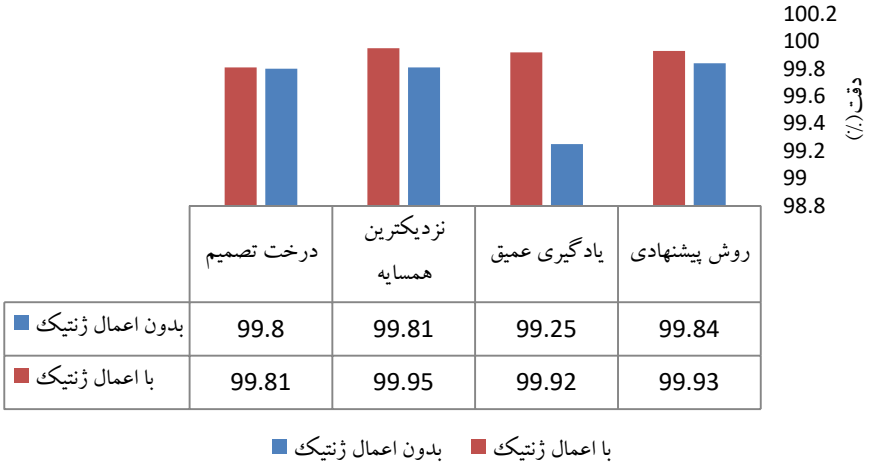
در جدول (۳) نتایج کلی الگوریتم‌های شبکه عصبی عمیق، درخت تصمیم، KNN و ترکیب این روش‌ها با اعمال الگوریتم ژنتیک و بدون الگوریتم ژنتیک آن نشان داده‌شده است

همان‌طور که از شکل بالا مشاهده می‌شود، نتایج دقت تشخیص و فراخوانی روش پیشنهادی به‌منظور جلوگیری از نفوذگران در شبکه در حالتی که با اعمال انتخاب ویژگی بوده است در مقایسه با حالتی که بدون اعمال الگوریتم انتخاب ویژگی ژنتیک بوده است بهتر عمل می‌کند.

در شکل زیر مقایسه نتایج روش پیشنهادی با اعمال الگوریتم انتخاب ویژگی ژنتیک نسبت به سایر روش‌ها و بدون اعمال ژنتیک نشان داده‌شده است

همان‌طور که از شکل بالا مشاهده می‌شود، نتایج روش پیشنهادی با اعمال الگوریتم انتخاب ویژگی ژنتیک برابر با ۹۹٫۹۳٪، روش یادگیری عمیق برابر با ۹۹٫۹۲٪، روش نزدیک‌ترین همسایه برابر با ۹۹٫۹۵٪ و روش درخت تصمیم برابر با ۹۹٫۸۱٪ است و بدون اعمال ژنتیک درروش پیشنهادی برابر با ۹۹٫۸۴٪، روش یادگیری عمیق برابر با ۹۹٫۲۵٪، روش نزدیک‌ترین همسایه برابر با ۹۹٫۸۱٪ و روش درخت تصمیم برابر با ۹۹٫۸٪ است؛ بنابراین با توجه به نتایج بالا روش

مقایسه نتایج روش پیشنهادی با اعمال الگوریتم انتخاب ویژگی ژنتیک
نسبت به سایر روشها و بدون اعمال ژنتیک



	روش پیشنهادی	یادگیری عمیق	نزدیکترین همسایه	درخت تصمیم
بدون اعمال ژنتیک	99.84	99.25	99.81	99.8
با اعمال ژنتیک	99.93	99.92	99.95	99.81

شکل ۱۰: مقایسه نتایج روش پیشنهادی با اعمال الگوریتم انتخاب ویژگی ژنتیک نسبت به سایر روش‌ها و بدون اعمال ژنتیک

پیشنهادی با اعمال ژنتیک در حدود ۰٫۰۹٪ نسبت به بدون اعمال ژنتیک بهبود داشته است و روش‌های یادگیری عمیق، نزدیک‌ترین همسایه و درخت تصمیم با اعمال ژنتیک به ترتیب در حدود ۰٫۶۷٪، ۰٫۱۴٪ و ۰٫۰۱٪ نسبت به انتخاب ویژگی بدون اعمال ژنتیک بهبود داشته‌اند.

۵-۴- مقایسه نتایج روش پیشنهادی با سایر روش‌ها

در این بخش به مقایسه نتایج روش پیشنهادی با اعمال الگوریتم انتخاب ویژگی ژنتیک و بدون اعمال این الگوریتم بر روی دیتاست مشابه و در شرایط یکسان با مقاله [۲۲] پرداخته می‌شود. در ادامه به تفسیر و مقایسه نتایج پرداخته می‌شود. در جدول زیر مقایسه نتایج روش پیشنهادی با سایر روش‌های مطرح‌شده در مقاله [۲۲] از جنبه میزان دقت تشخیص نشان داده‌شده است.

در شکل ۱۱ نیز مقایسه دقت تشخیص نفوذ گران درروش پیشنهادی نسبت به سایر روش‌های مطرح‌شده در مقاله [۲۲] بر روی دیتاست KDDCup99 نشان داده‌شده است

همان‌طور که از شکل بالا مشاهده می‌شود، میزان دقت تشخیص نفوذگران درروش پیشنهادی با استفاده از دیتاست KDDCup99 برابر با ۹۹٫۹۳٪، روش Grid SVM برابر با ۹۲٫۷۵٪، روش PSO SVM برابر با ۹۴٫۳۴٪، روش GA SVMبرابر با ۹۵٫۸۹٪، روش جنگل تصادفی برابر با ۹۵٫۹۸٪، روش شبکه بیزین برابر با ۹۵٫۳۳٪ و روش HG-GA SVM برابر با ۹۶٫۷۲٪ است.

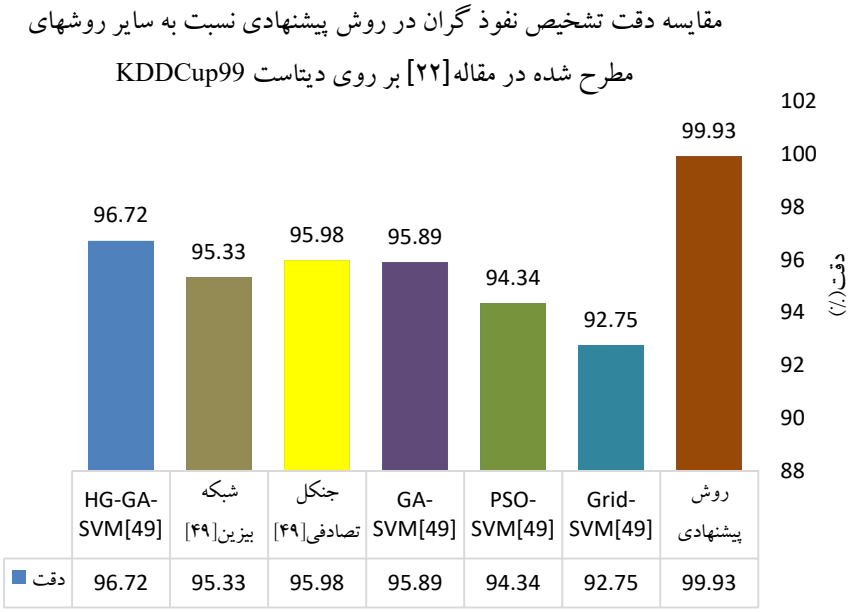
با این تفاسیر میزان بهبود دقت روش پیشنهادی در مقایسه به سایر روش‌های Grid SVM، PSO SVM، GA SVM، جنگل تصادفی، شبکه بیزین و HG-GA SVM به ترتیب برابر با ۷٫۱۸٪، ۵٫۵۹٪، ۴٫۰۴٪، ۳٫۹۵٪، ۴٫۶٪ و ۳٫۲۱٪ است.

یکی از معیارهای مهم دیگر نرخ تشخیص درست است که این معیار میزان درستی روش پیشنهادی را مطرح می‌کند. در شکل زیر مقایسه نرخ تشخیص درست درروش پیشنهادی با اعمال انتخاب ویژگی نسبت به سایر روش‌های مطرح‌شده در [۲۲] نشان داده‌شده است.

نتایج نشان داده‌شده در شکل ۱۲ تأییدی بر اعتبارسنجی و دقت قابل‌توجه روش پیشنهادی

جدول ۴- مقایسه نتایج روش پیشنهادی با سایر روش‌های مطرح‌شده در مقاله [۲۲] از جنبه میزان دقت تشخیص

	بدون اعمال انتخاب ویژگی	با اعمال انتخاب ویژگی
	دقت	دقت
روش پیشنهادی	99.84	99.93
Grid-SVM[49]	90.13	92.75
PSO-SVM[49]	94.54	94.34
GA-SVM[49]	95.33	95.89
جنگل تصادفی[۴۹]	95.98	95.98
شبکه بیزین[۴۹]	95.33	95.33
HG-GA-SVM[49]	95.82	96.72



شکل ۱۱- مقایسه دقت تشخیص نفوذ گران درروش پیشنهادی نسبت به سایر روش‌های مطرح‌شده در مقاله [۲۲] بر روی دیتاست KDDCup99

نسبت به دیگر روش‌های نوین در مطالعات پیشین است.

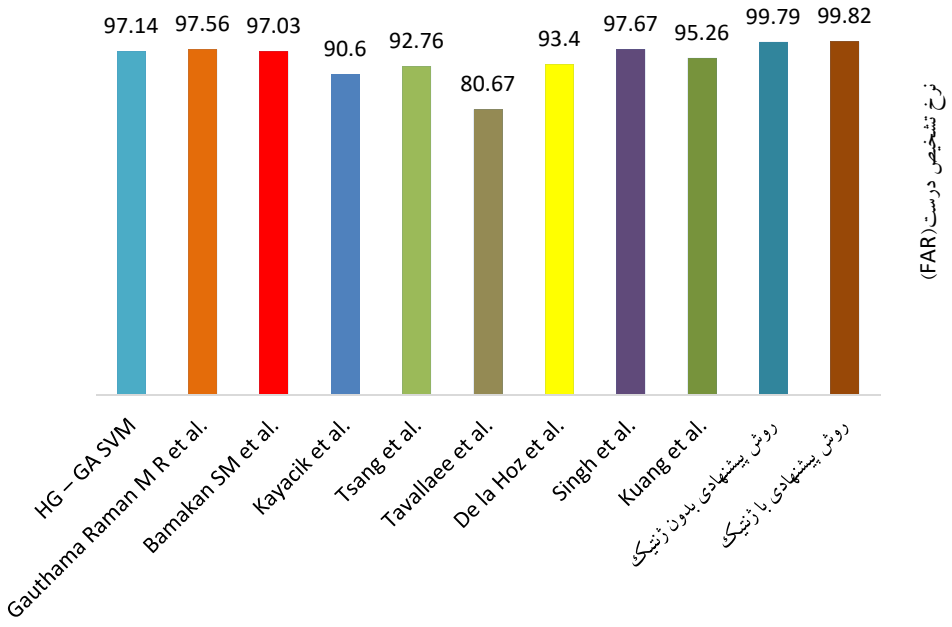
همان‌طور که از شکل بالا مشاهده می‌شود، میزان بهبود نرخ تشخیص درست درروش پیشنهادی با اعمال انتخاب ویژگی نسبت به روش پیشنهادی بدون ژنتیک و روش‌های Kuang et al ، Singh et al، De la Hoz et al، Tavallaee et al، Tsang et al، Kayacik et al، Bamakan SM et al، Gauthama Raman MR et al و HG-GA SVM به ترتیب برابر با ۰,۰۳٪، ۴,۵۶٪، ۲,۱۵٪، ۶,۴۲٪، ۱۹,۱۵٪، ۷,۰۶٪، ۹,۲۲٪، ۲,۷۹٪، ۲,۲۶٪ و ۲,۶۸٪ است.

۶- نتیجه‌گیری

در این مقاله برای تشخیص حملات در سیستم بانکداری الکترونیک، از ترکیب الگوریتم انتخاب ویژگی ژنتیک و روش‌های یادگیری ماشین ازجمله الگوریتم درخت تصمیم، شبکه عصبی عمیق و KNN به‌صورت تلفیقی استفاده‌شده است.

برای اعتبار سنجی راهکار ارائه‌شده، نتایج حاصل با سایر روش‌ها ازجمله روش‌های یادگیری ماشین و ترکیبی با سایر روش‌های بهینه‌سازی مورد مقایسه و ارزیابی قرارگرفته است. در این پژوهش از ۱۰٪ مجموعه داده KDD Cup 99 برای شبیه‌سازی استفاده‌شده است که ابتدا در مرحله

مقایسه نرخ تشخیص درست در روش پیشنهادی با اعمال انتخاب ویژگی نسبت به سایر روشهای مطرح شده در [۲۲]



شکل ۱۲- مقایسه نرخ تشخیص درست درروش پیشنهادی با اعمال انتخاب ویژگی نسبت به سایر روش‌های مطرح‌شده در [۲۲].

پیش‌پردازش داده‌ها، مقادیر کلیه مشخصه‌ها به اعداد تبدیل و همچنین مقادیر مشخصه خروجی به دو مقدار صفر و یک تغییر داده‌شده است. نتایج حاصل از پژوهش نشان از دقت بالای راهکار ارائه‌شده برای تشخیص نفوذگران نسبت به سایر روش‌های اخیر در حدود ۵٪ است.

۷- منابع

۱-"شرکت بهین راهکار توسعه پیشرو," ۱۳۹۹. [درون‌خطی :Available] . http://www.behinrahkar.com.

۲- "پروژه معماری و طراحی مرکز عملیات امنیت بومی مبتنی بر راهکارهای متن‌باز," ۱۳۸۸. [درون‌خطی .] Available: http://www.ic4i.ir.

3. E. ElFgee and A. Arara, "Technical Requirements of New Framework for GPRS Security Protocol Mobile Banking Application," Procedia Computer Science, p. Procedia Computer Science, 2014.

4. S. Bojjagani and V. N. Sastry, "Stamba: Security testing for Android mobile banking apps," In Advances in Signal Processing and Intelligent Recognition Systems, pp. pp. 671-683, 2016.

5. R. K. Rehiman and S. Veni, "A secure authentication infrastructure for IoT Enabled smart mobile devices–An Initial Prototype," Indian Journal of Science and Technology, p. 9(9), 2016.

6. N. Thompson, L. T. McGill and X. Wang, ""Security begins at home": Determinants of home computer and mobile device security behavior," computers & security, pp. 70, 376-391, 2017.

7. S. Bhatnagar, Y. Malik and S. Butakov, "Analysing Data Security Requirements of Android Mobile Banking Application," in In International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, 2018.

8. S. Chen,. Meng, T. Su, L. Fan, Y. Xue, Y. Liu and S. Hao, "AUSERA: Large-Scale Automated Security Risk Assessment of Global Mobile Banking Apps," arXiv preprint arXiv, p. 1805.05236, 2018.

9. H. K. Yeh, "A secure transaction scheme with certificateless cryptographic primitives for IoT-based mobile payments," IEEE Systems Journal, pp. 12(2), 2027-2038, 2018.

10. D. Saurabh, Y. Qiang and S. Srinivas, "A Machine Learning Based Intrusion Detection Scheme for Data Fusion in Mobile Clouds Involving Heterogeneous Client Networks," Information Fusion, 2018.

11. S. Sakr, A. Liu, M. D. Batista and M. Alomari, "A survey of large scale data management approaches in cloud environments," IEEE Communications Surveys

& Tutorials, pp. 13(3), 311-336, 2011.
12. 2021. [Online]. Available: https://www.tutorialspoint.com/cryptography/advanced_encryption_standard.htm.
13. H. M. Mokhtar, O. Ossama and M. E.El-Sharkawi, "An extended k-means technique for clustering moving objects," Egyptian Informatics Journal, pp. 45-51, 2011.
14. S. N. Sivanandam and S. N. Deepa, "Genetic algorithm optimization problems," In Introduction to Genetic Algorithms, pp. pp. 165-209, 2008.
15. D. S. Weile and E. Michielssen, "Genetic algorithm optimization applied to electromagnetics: A review," IEEE Transactions on Antennas and Propagation, pp. 45(3), 343-353, 1997.
16. K. Indira and S. Kanmani, "Association rule mining using genetic algorithm: The role of estimation parameters," in International Conference on Advances in Computing and Communications, Berlin, Heidelberg, 2011.
17. D. E. Goldberg, Genetic Algorithm in Search, Optimization & Machine Learning, New York: Addison-Wesely, 1989.
18. N. Jain and V. Srivastava, "Data Mining techniques: A survey paper.IJRET," International Journal of Research in Engineering and Technology, pp. pp.19-23, 2013.
19. T. Shang, X. Xia and J. Zheng, "MIME-KNN: Improve KNN Classifier Performance Include ردهبندی Accuracy and Time Consumption," DEStech Transactions on Computer, 2018.
20. S. Yang, H. Jian, Z. Ding, Z. Hongyuan and G. C. Lee, IKNN: Informative K-Nearest Neighbor Pattern Classification, Berlin Heidelberg: Springer-Verlag, 2007.
21. Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," nature, pp. 521(7553), 436, 2015.
22. M. G. Raman, N. Somu, K. Kirthivasan, R. Liscano and V. S. Sriram, "An efficient intrusion detection system based on hypergraph-Genetic algorithm for parameter optimization and feature selection in support vector machine," Knowledge-Based Systems, pp. 134, 1-12, 2017.
۲۳- ه. مینتربرگ، ب. آلستراند و ژ. لمپل، جنگل استراتژی (کارآفرینی در قالب یک مکتب)، تهران: جاجرمی، ۱۳۸۴ .
24. B. Schwenker and T. Wulf, Scenario-based Strategic Planning: Developing Strategies in an Uncertain World, Springer, 2013.
۲۵- ع. آذر، آمار و کاربرد آن در مدیریت، تهران: سمت، ۱۳۹۰ .
26. M. Portet, Competitive strategy: Techniques for analyzing industries and competitors, New York: Simon and Schuster, 2008.
27. G. Johnson, K. Scholes and R. Whittington, Exploring Corporate Strategy, 8th ed., Londen: Financial Times Prentice Hall, 2008.