هفتمین کنفرانس ملی مباحث نوین در کامپیوتر و فناوری اطلاعات
7th National Conference on Advanced Topics in Computer and Information Technology
۲ آذر ماه ۱۴۰۲

استفاده از هوش مصنوعی جهت تشخیص انواع بدافزارها در سیستم‌های رایانه‌ای

شکوفه جهان بزرگی مطلق - نسیم حمیدیان

چکیده : در سال‌های اخیر یکی از مهمترین چالش‌های امنیت اطلاعات و شبکه‌های ارتباطی افزایش روزافزون انواع بدافزارها و به دنبال آن یافتن راه‌های مناسب جهت حفاظت سیستم‌ها در مقابل آنهاست که از مهمترین دغدغه‌های برنامه‌نویسان و متخصصین امنیت اطلاعات شناخت به موقع و یافتن راه‌های مقابله با اثرات مخرب این‌گونه بدافزارها می‌باشد که در مقابله با این روند شرکت‌های ضدبدافزاری به طور جدی به دنبال راه‌های بهتری برای شناسایی بدافزار هستند در این راستا ایجاد روشی که بتواند به‌صورت کارا به شناسایی و جلوگیری از نفوذ بدافزارها بپردازد همواره مورد نیاز خواهد بود. استفاده از هوش مصنوعی یکی از روش‌هایی است که محققان توانسته‌اند در سال‌های اخیر با استفاده از آن به تشخیص بدافزارها با دقت‌های مختلفی بپردازند و در این راه به پیشرفت‌هایی در تشخیص بدافزار دست یافته‌اند. در این مقاله سعی نمودیم جهت تشخیص انواع بدافزارها در سیستم‌های رایانه‌ای از الگوریتم‌های چندگانه شبکه عصبی مصنوعی، ماشین بردار پشتیبان و درخت تصمیم مبتنی بر رای‌گیری حداکثری استفاده نماییم. نتایج بدست آمده حاکی از آن است که روش پیشنهادی با دقت ۹۵.۱۷ درصد برای داده‌های آموزشی و دقت ۹۳.۷۴ درصد برای داده‌های آزمایشی قادر به تشخیص انواع بدافزارها در سیستم‌های رایانه‌ای می‌باشد.

کلمات کلیدی: تشخیص- بدافزار - سیستم‌های رایانه ای – هوش مصنوعی

۱ - بدافزار

بدافزار برنامه‌ای می‌باشد که به طور عمده برای انجام فعالیت‌های مخربی از جمله سرقت اطلاعات کاربر، جاسوسی اینترنتی و غیره طراحی گردیده است. پویایی رفتاری بدافزار به عوامل مختلفی همچون ماهیت حمله، فناوری پیشرفته و افزایش سریع آسیب‌پذیری وابسته خواهد بود. امروزه علاوه بر افزایش استفاده از دستگاه‌های دیجیتال و اینترنت، حمله‌های مخرب نیز با رشد سریعی همراه است. افزایش چشمگیر در نرخ ایجاد نرم‌افزارهای مخرب جدید در طی پنج سال گذشته، سبب مطرح شدن روش‌های تشخیص و شناسایی بدافزار به عنوان یک مسئله مهم تحقیقی شده است [۱]. رشد سریع بدافزار سبب بوجود آمدن تهدیدهای بسیار زیادی در حوزه امنیت اطلاعات شده است. بنابراین مراکز دفاع سایبری که همانند مرزهای یک کشور هستند از اهمیت زیادی برخوردار بوده و در کنار فضای مجازی می‌توانند مورد تهدید و هجوم قرار بگیرند. نرم‌افزارهای مخرب در طول زمان از یک مزاحمت غیرجدی تبدیل به یک تهدید امنیتی جدی شده‌اند که این امر سبب شده است توجه محققان امنیتی در سراسر جهان به این موضوع جلب شود. بنابراین به منظور تعیین رفتار، حالت وخیم بودن و عملیات انواع نرم‌افزارهای مخرب؛ باید مورد تجزیه و تحلیل قرار بگیرند [۲].

جدول ۱: تاریخچه ایجاد برخی بدافزارها در جهان

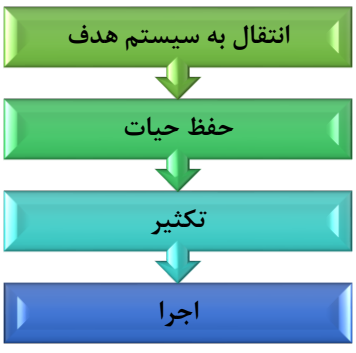
نام بدافزار	کشور تولیدکننده	سال انتشار	عملکرد بدافزار تولید شده
Brain	پاکستان	۱۹۸۶	اولین ویروس برای کامپیوترهای شخصی
Stoned	نیوزلند	۱۹۸۷	توسط یک دانش‌آموز دبیرستانی نوشته شد
Form	سوئیس	۱۹۹۰	یکی از بزرگترین موارد پخش ویروس در سراسر جهان
Michelangelo	نیوزلند	۱۹۹۱	نخستین ویروس کامپیوتری که وارد ادبیات رسانه‌ای شد
VCL	آمریکا	۱۹۹۲	تولد اتوماتیک ویروس با یک ظاهر ساده گرافیکی
Monkey	کانادا	۱۹۹۴	برنامه با قابلیت مخفی‌سازی خودکار
Concept	آمریکا	۱۹۹۵	اولین ویروسی که فایل‌های مایکروسافت ورد را آلوده کرد
Happy99	نامشخص	۱۹۹۹	اولین ویروس ایمیلی
Melissa	آمریکا	۱۹۹۹	ظاهرا ملیسا نام یک رقصنده است
Code Red	نامشخص	۲۰۰۱	شیوع اولین کرم بدون هیچ نوع دخالت کاربر
Love Letter	فیلیپین	۲۰۰۰	با attachment ایمیل یکی از بزرگترین موارد شیوع و تخریب
Slammer	نامشخص	۲۰۰۳	خودپردازهای Bank of America و ۹۱۱ سیاتل را از کارانداخت
Sobig	نامشخص	۲۰۰۳	در کمتر از چند ساعت میلیون‌ها کامپیوتر را آلوده کرد
Fizzer	نامشخص	۲۰۰۳	اولین ویروس با هدف کسب درآمد با ارسال اسپم
Cabir	فیلیپین	۲۰۰۳	اولین کرم روی تلفن‌های موبایل
MyDoom	روسیه	۲۰۰۴	تکثیر از طریق ایمیل و شبکه‌های محبوب p2p مشخصاKazaa
Sasser	آلمان	۲۰۰۴	شبکه‌ها را از استرالیا تا هنگ کنگ و انگلستان از کار انداخت

جدول۲: مقایسه بدافزارها

پارامترهای مقایسه	خانواده بدافزار	نرم‌افزارهای جاسوسی	نرم‌افزارهای تبلیغاتی	کرم‌ها	درب پشتی	تروجان	اسپیفرها	اسپم	بات نت	پمپ متقلبی	کرم	ویروس
تکنیک‌های تولید	الگو	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	مبهم‌سازی	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	چندریختی	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ابزار خاص	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
محیط اجرایی	شبکه	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×
	اجرا از راه دور ازطریق وب	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×
	کامپیوتر	×	×	×	×	×	×	×	×	×	✓	✓
رسانه انتشار	شبکه	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	حافظه‌های قابل حمل	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	دانلودهای اینترنتی	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
اثرات منفی	نقض حریم شخصی	✓	×	✓	×	✓	✓	×	×	×	×	×
	ایجاد مشکلات برای کاربران	×	✓	×	×	×	×	✓	×	×	×	×
	اختلال در کارکرد سرویس‌ها	×	×	×	✓	×	×	✓	✓	✓	✓	✓
	دستکاری داده‌ها	×	×	×	✓	×	×	✓	✓	✓	×	✓

۲- شیوه عملکرد بدافزارها

بطور کلی چهار مرحله اصلی در چرخه عمر بدافزارها می‌توان در نظر گرفت که عبارتند از [۳]:



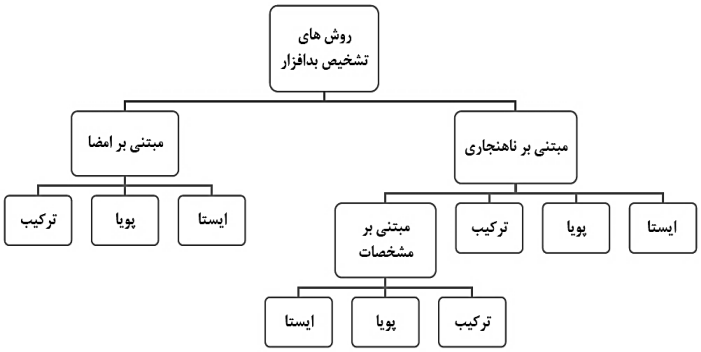
شکل ۱: چرخه عمر بد افزارها

- **انتقال به سیستم هدف:** بدافزارها به شیوه‌های مختلفی به سیستم‌های کامپیوتری منتقل می‌شوند با اجرای یک فایل آلوده از یک ذخیره‌ساز اطلاعات، دانلود یک فایل از اینترنت یا سایر شبکه‌های کامپیوتری، بهره‌گیری از آسیب‌پذیری‌های یک سیستم کامپیوتری و غیره. برخی از این انتقال‌ها به صورت هدفمند می‌باشد، به طور مثال به منظور دستیابی به اطلاعات یا از کار انداختن سرویس یک شرکت خاص بوده و جنبه عمومی ندارد. دسته‌ای دیگر ممکن است همه کاربران اینترنت را مورد هدف قرار داده و به صورت عمومی فعالیت نماید.
- **حفظ حیات:** بدافزارها پس از انتقال به سیستم هدف برای دستیابی به اهداف خود باید به گونه‌ای در سیستم پنهان شده تا در زمان مناسب و با فراهم شدن شرایط لازم نظیر دسترسی به منبع خاص از سیستم یا بارگذاری در حافظه Ram فعالیت‌های مورد نظر خود را انجام دهند که ممکن است در یکی از فایل‌های اجرایی سیستم ادغام شوند یا در مکانی از حافظه پنهان گردند.
- **تکثیر:** هر کدام از بدافزارها برای تکثیر نیاز به فراهم شدن شرایط خاصی دارند. نوع تکثیر و هدف آن نیز متفاوت می‌باشد. برای مثال wormها پس از استقرار در حافظه RAM شروع به تکثیر خود می‌کنند، که هدفشان از این کار معمولا کند شدن سرعت سیستم است. ویروس‌ها نیز خود را تکثیر می‌کنند که یکی از اهداف آن‌ها آلوده کردن تعداد بیشتری از فایل‌های سیستم می‌باشد که با اجرای فایل‌های آلوده شده عملکرد سیستم توسط ویروس در معرض آسیب قرار می‌گیرد. البته ممکن است برخی از انواع بدافزارها بدون تکثیر عمل کنند و فقط یک فایل خاص را آلوده نمایند. تروجان‌ها معمولا تکثیر نمی‌شوند و از این دسته می باشند.
- **اجرا:** پس از مستقر شدن بدافزارها باید به گونه‌ای به حافظه اصلی سیستم منتقل شوند تا بتوانند دستورات خود را اجرا کنند. استقرار بدافزارها از قبیل ویروس‌ها در برنامه‌های کاربردی سیستم، یکی از روش‌های متداول می‌باشد که بیشترین آسیب را متوجه کاربران نموده است. این گونه برنامه‌ها به ظاهر مفید اما در اصل مخرب می‌باشند. به این منظور ویروس، کدهای مخرب خود را به برنامه‌های کاربردی Append می‌کند که پس از اجرای برنامه توسط کاربر، در حین اجرا این کدهای مخرب نیز اجرا می‌شوند.

۳- روش‌های تشخیص بدافزار

روش‌های تشخیص بدافزار برای تشخیص بدافزارها مورد استفاده قرار گرفته و از آلوده شدن سیستم‌های کامپیوتری و از دست رفتن اطلاعات و سازش سیستم جلوگیری می‌کنند. بطور کلی تشخیص بدافزار به دو

دسته کلی تشخیص مبتنی بر امضا[1] و تشخیص مبتنی بر ناهنجاری[2] تقسیم می‌شود. هریک از روش‌های تشخیص بدافزار خود شامل روش‌های پویا[3]، روش‌های ایستا[4] و روش‌های ترکیب[5] هستند. در شکل ۲-۲ انواع روش‌های تشخیص بدافزار نمایش داده شده است [۴،۵].



شکل ۲: طبقه‌بندی انواع روش‌های تشخیص بدافزار [۶]

جدول ۱: مزایا و معایب روش‌های مبتنی بر امضا و ناهنجاری

تجزیه و تحلیل	مزایا	معایب
مبتنی بر ناهنجاری	زمان اسکن بسیار پایین است	ناتوان در تشخیص بدافزارهای ناشناخته
	میزان نرخ مثبت غلط پایین است	توانایی اندک در شناسایی موارد مبهم و گیج‌کننده
مبتنی بر امضا	ارائه بهترین نتایج در تشخیص بدافزارهای چندریختی از ویروس‌های بسته‌بندی شده نیست	قادر به تشخیص بسیاری

۴- هوش مصنوعی

سیستم‌های هوش مصنوعی (AI) سیستم‌های نرم‌افزارها و حتی گاه سخت‌افزارها هستند که توسط انسان طراحی شده‌اند تا در صورتی که یک تکلیف پیچیده به عهده آنان قرار گیرد آنان بصورت فیزیکی یا در ابعاد دیجیتالی با درک محیط اطراف خود از طریق به دست آوردن داده‌ها و تحلیل و تفسیر داده‌های به دست آورده ساختاریافته یا ساختارنیافته[6]، استدلال کردن براساس یافته‌ها یا با تجزیه و تحلیل اطلاعاتی که از این داده‌ها به دست آورده‌اند عمل کنند و برای انجام دادن بهترین عمل با هدف انجام دادن تکلیف اولیه تصمیم بگیرند. سیستم‌های هوش مصنوعی می‌توانند قواعد نمادین[7] یا مدل‌های اعدادی[8] را یاد بگیرند. آنها همچنین می‌توانند رفتار خود را براساس تحلیل تأثیر اعمال قبلی خود بر محیط وفق دهند. در سال‌های اخیر در پی پیشرفت‌هایی که در زمینه رایانه و هوش مصنوعی و همچنین کشف روابط آشوبی در سری‌های زمانی غیرخطی پدید آمد، فعالیت‌هایی در جهت پیش‌بینی قیمت در بورس اوراق بهادار در کشورهای مختلف انجام شد. تکنیک‌های هوش مصنوعی که شامل شبکه‌های عصبی، داده کاوی، الگوریتم ژنتیک و منطق فازی است نتایج موفقیت‌آمیزی در زمینه حل مسایل پیچیده به دست آورده‌اند[۷].

هوش مصنوعی به شکل ساده آن بیش از ده‌ها سال است که وجود داشته است. بطور عادی یادگیری ماشین در قالب الگوریتم‌های ساده که دستورهایی هستند که برنامه‌نویس آنها را طراحی کرده مورد استفاده قرار می‌گرفت. در ابتدا کارکرد هوش مصنوعی به کارهای محدودی با داده‌های محدود خلاصه می‌شد. امروزه انسان‌ها با استفاده از اینترنت و دستگاه‌هایی که به نحوی به اینترنت متصل‌اند[9] در حال تولید داده‌ها در حجمی گسترده‌اند. در نتیجه انسان‌ها در آغاز قادر به تحلیل حجم سنگینی از داده‌ها در زمان بهینه نبودند. بنابراین محققان دست به طراحی برنامه‌هایی زدند که به جای آنکه برنامه‌نویس به تحلیل داده‌ها و نتیجه‌گیری بپردازد نرم‌افزار طراحی شده چنین تحلیلی را انجام دهد. امروزه بیش از پانزده رویکرد برای استفاده و طراحی یادگیری ماشین وجود دارد که هر یک از یک نوع الگوریتم استفاده می‌کنند[۸]. هدف اصلی یک الگوریتم دسته‌بندی به حداکثر رساندن دقت پیش‌بینی است و برای این منظور الگوریتم‌های گوناگونی ایجاد شده است که برخی از مهم‌ترین آنها به شرح زیر است[۹].

[1] signature-based methods
[2] anomaly-based methods
[3] dynamic methods
[4] satic methods
[5] hybrid methods

[6] structured or unstructured data
[7] symbolic rules
[8] numeric model
[9] Internet of things

۴-۱- درخت تصمیم[10]

درخت تصمیم یکی از معروف‌ترین روش‌های دسته‌بندی داده‌ها در داده کاوی است که برای کمک به تصمیم‌گیری ساخته می‌شود. از این ساختار برای تقسیم مجموعه‌ای بزرگ از داده‌های جمع‌آوری شده به مجموعه‌های کوچک‌تر استفاده می‌شود، این عمل به واسطه یک مجموعه از قوانین ساده تصمیم‌گیری صورت می‌گیرد. در هر تقسیم‌بندی متوالی، اعضای مجموعه‌های حاصل بیش از پیش به یکدیگر شبیه می‌شوند. درخت تصمیم یک تقسیم‌بندی کننده بازگشتی فضای نمونه است. این مدل از چند گره و یک ریشه تشکیل شده است و همه گره‌ها به جز ریشه دقیقا یک لبه ورودی دارند. یک درخت تصمیم با یک گره اولیه شروع می‌شود که پس از آن پیامدهای احتمالی به صورت شاخه‌هایی از آن منشعب شده و هر کدام از آن پیامدها به گره‌های دیگری منجر شده که آنها نیز شاخه‌هایی از احتمالات دیگر را ایجاد می‌کنند. به گره‌های بدون خروجی برگ گفته می‌شود (به عنوان گره‌های انتهایی یا تصمیم‌گیری نیز شناخته می‌شود). در یک درخت تصمیم، هر گره داخلی فضای نمونه را به دو یا چند فضای فرعی تقسیم می‌کند که یک تابع گسسته خاص از مقادیر ویژگی‌های ورودی است[۱۰].

۴-۲-ماشین‌های بردار پشتیبان[11]

ماشین‌های بردار پشتیبان یک روش بسیار موثر برای رگرسیون، دسته‌بندی و تشخیص الگوی کلی هستند اما به طور گسترده در اهداف دسته‌بندی استفاده می‌شوند. این روش با نام SVM نیز شناخته می‌شوند که مخفف Support Vector Machine می‌باشد. هدف از اجرای الگوریتم ماشین‌های بردار پشتیبان پیدا کردن یک هایپرپلین یا زیرصفحه (hyperplane) در یک فضای N بعدی (N برابر است با تعداد ویژگی‌ها) است که به طور مشخص نقاط داده را دسته‌بندی می‌کند. برای جدا کردن دو دسته از نقاط داده می‌توان زیرصفحه‌های زیادی را انتخاب کرد اما در این روش هدف ما یافتن صفحه‌ای است که دارای حداکثر حاشیه باشد، به این معنی که باید بین نقاط داده هر دسته با دسته دیگر حداکثر فاصله وجود داشته باشد تا نقاط آینده را بتوان را با اطمینان بیشتری دسته‌بندی کرد. زیرصفحه‌ها مرزهای تصمیم‌گیری هستند که به دسته‌بندی نقاط داده کمک می‌کنند. نقاط داده‌ای که در دو طرف زیرصفحه‌ها قرار دارند را می‌توان به دسته‌های مختلف نسبت داد. همچنین ابعاد زیرصفحه‌ها به تعداد ویژگی‌ها بستگی دارد. به عنوان مثال اگر دو ویژگی ورودی داشته باشیم زیرصفحه فقط یک خط دارد و اگر سه ویژگی ورودی داشته باشیم آنگاه زیرصفحه به به صفحه‌ای دو بعدی تبدیل می‌شود[۱۱].

۴-۳-شبکه عصبی مصنوعی[12]

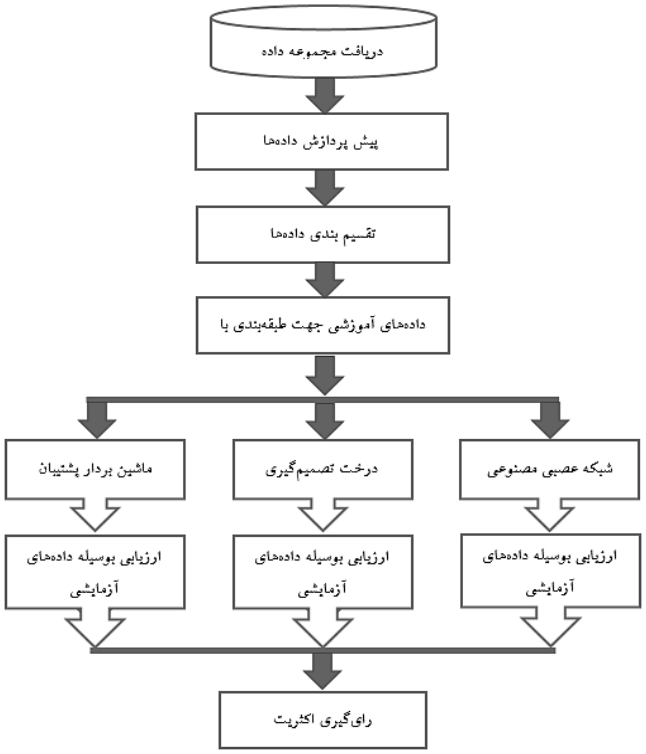
شبکه عصبی مصنوعی یکی از روش‌های دسته‌بندی داده‌ها در داده کاوی است که در آن مدل یاد گرفته شده به صورت مجموعه‌ای از گره‌های به هم متصل به همراه ارتباطات وزن‌دار آنها نمایش داده می‌شود. شبکه‌های عصبی برای تشخیص روابط در داده‌های با حجم بالا از عملکرد مغز انسان الگو می‌گیرند به همین دلیل در نام آنها از واژه نورون(Neuron) استفاده شده است و دارای دو نوع ارگانیک (Organic) و مصنوعی (Artificial) هستند. شبکه‌های عصبی در فرآیند آموزش خود اطلاعات مربوطه را در شبکه ذخیره می‌کنند. این شبکه توانایی یادگیری دارد و می‌تواند با توجه به اطلاعات اولیه چیزی را فرا بگیرد و یا براساس آموخته‌های خود تصمیم‌گیری کند. شبکه‌های عصبی می‌توانند با تغییر ورودی سازگار شوند به همین دلیل بدون نیاز به طراحی مجدد معیارهای خروجی می‌توانند بهترین نتیجه ممکن را تولید کنند. همان طور که گفتیم یک شبکه عصبی مانند شبکه عصبی مغز انسان عمل می‌کند که نورون در آن یک تابع ریاضی است و اطلاعات را با توجه به معماری خاص خود جمع‌آوری و طبقه‌بندی می‌کند. این شبکه شباهت زیادی به روش‌های آماری مانند برازش منحنی و تحلیل رگرسیون دارد. یک شبکه عصبی شامل چند لایه از گره‌های به هم پیوسته است. هر گره یک گیرنده است و شبیه به یک رگرسیون خطی چندگانه است. شبکه‌های عصبی که ریشه در هوش مصنوعی دارند به سرعت در توسعه سیستم‌های تجاری محبوبیت پیدا کردند و امروزه از آنها در نرم‌افزارهای مختلف مالی، تحقیقات بازار، پیش‌بینی و بازاریابی، کشف تقلب و ارزیابی ریسک استفاده می‌شود[۱۲].

۵- روش پیشنهادی

در این تحقیق جهت تشخیص انواع بدافزارها در سیستم‌های رایانه‌ای از الگوریتم‌های هوش مصنوعی(شبکه عصبی مصنوعی، ماشین بردار

[10] Decision Tree
[11] Support Vector Machines
[12] Artificial Neural Networks

پشتیبان و درخت تصمیم) استفاده نمودیم. روش پیشنهادی در این تحقیق قادر است که فایل‌های آلوده به انواع مختلف بدافزارها را بصورت کاملاً قابل اعتماد و کارآمد شناسایی کند.



شکل ۳: دورنمای روش پیشنهادی

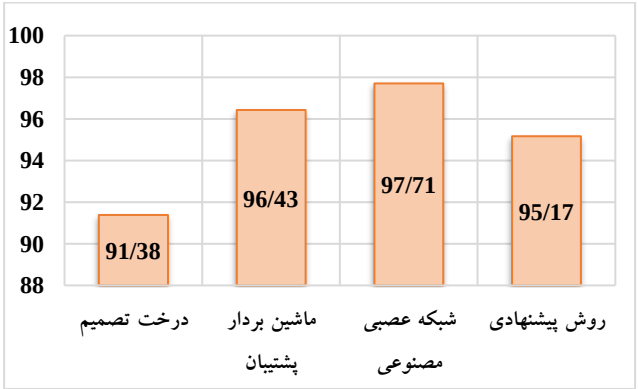
برای این منظور در ابتدا به جمع‌آوری مجموعه داده مورد نظر پرداخته شد. بدین منظور داده‌های مورد نظر از سایت Kaggle[۱۳] گردآوری گردید. مجموعه داده دریافت شده شامل ۱۰۰۰۰۰ رکورد یا فایل می‌باشد که هر فایل دارای ۳۲ ویژگی است. در این مجموعه داده[۱۴] ۳۱ ویژگی به عنوان ویژگی‌های اصلی بوده و یک ویژگی آن به عنوان برچسب[۱۵] و تعیین کننده نوع فایل از جهت سالم و یا آلوده بودن به بدافزار می‌باشد. سپس داده‌ها را مورد نرمالسازی قرار دادیم در واقع نرمال‌سازی روشی است که داده‌ها در زمانی که در یک دامنه قرار ندارند را در دامنه مشابه قرار می‌دهد. برای این منظور با استفاده از روش نرمالسازی Max-Min داده‌های موجود در مجموعه داده را در بازه بین صفر و یک نرمالسازی کردیم که در رابطه ۱-۳ نشان داده شده است.

$$\acute{V} = \frac{V - min(V)}{max(V) - min(V)} \qquad (۱-۳)$$

که در این رابطه V مقدار اصلی ویژگی مربوط به رکورد یا فایلی که قرار است نرمال شود می‌باشد، $\acute{V}$ مقدار نرمال‌سازی شده ویژگی مربوط به رکورد مورد نظر است، max(V) حد بالای ویژگی و min(V) حد پایین مقدار ویژگی در بین کل رکوردهای مجموعه داده می‌باشد. در ادامه با استفاده از روش هولداوت داده‌ها به دو مجموعه داده آموزشی و آزمایشی با نسبت ۷۰ به ۳۰ درصد تقسیم شدند. پس از تقسیم‌بندی داده‌ها به ایجاد و ساخت الگوریتم‌های مورد استفاده پرداختیم و در ادامه به آموزش و ارزیابی هر یک از طبقه‌بندهای مورد نظر به تعداد ۱۰۰۱ مرتبه پرداخته و پس از رای‌گیری حداکثری از برچسب‌های هرکدام از طبقه‌بندهای ذکر شده به ارزیابی دقت آنها پرداخته و در نهایت از نتایج هرکدام میانگین‌گیری نموده و پاسخ بدست آمده بعنوان پاسخ نهایی در نظر گرفته شد.

۶- ارزیابی نتایج بدست آمده

در این بخش دقت بدست آمده از هرکدام از طبقه‌بندهای مورد استفاده با اعمال رای‌گیری حداکثری بوسیله داده‌های آموزشی و آزمایشی مورد بحث و بررسی قرار گرفته خواهد شد که در شکل ۴ و ۵ نشان داده شده است.

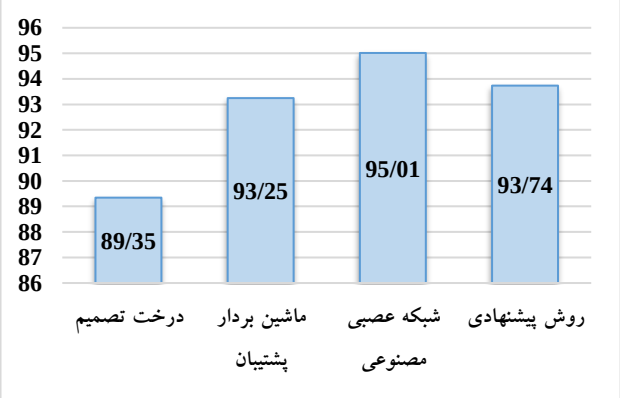


شکل ۴: مقایسه دقت روش پیشنهادی و طبقه‌بندهای مورد استفاده

با توجه به نتایج بدست آمده بوسیله داده‌های آموزشی که در شکل ۴ آمده است، مشاهده می‌کنیم که از میان طبقه‌بندهای مورد استفاده

[13] www.kaggle.com/nsaravana/malware-detection
[14] Dataset
[15] label

درخت تصمیم با دقت ۹۱.۳۸ درصد و طبقه‌بند شبکه عصبی مصنوعی با دقت ۹۷.۷۱ درصد دارای کمترین و بیشترین دقت بدست آمده هستند که اختلاف دقتی در حدود ۶.۳۳ درصد بین این دو طبقه‌بند وجود دارد. اما دقت در روش پیشنهادی برابر با ۹۵.۱۷ درصد می‌باشد که در واقع حد واسط بین دقت هرکدام از طبقه‌بندهای مورد نظر بوده و به عنوان یک نتیجه قابل قبول می‌توان از روش پیشنهادی در تشخیص انواع بدافزارها در سیستم‌های رایانه‌ای استفاده نمود.



شکل ۵: مقایسه دقت روش پیشنهادی و طبقه‌بندهای مورد استفاده

با توجه به نتایج بدست آمده بوسیله داده‌های آزمایشی که در شکل ۵ آمده است، مشاهده می‌کنیم که از میان طبقه‌بندهای مورد استفاده درخت تصمیم با دقت ۸۹.۳۵ درصد و طبقه‌بند شبکه عصبی مصنوعی با دقت ۹۵.۰۱ درصد دارای کمترین و بیشترین دقت بدست آمده هستند که اختلاف دقتی در حدود ۵.۶۶ درصد بین این دو طبقه‌بند وجود دارد. اما دقت در روش پیشنهادی برابر با ۹۳.۷۴ درصد می‌باشد که در واقع حد واسط بین دقت هر کدام از طبقه‌بندهای مورد نظر بوده و به عنوان یک نتیجه قابل قبول می‌توان از روش پیشنهادی در تشخیص انواع بدافزارها در سیستم‌های رایانه‌ای استفاده نمود.

۷– نتیجه‌گیری

امروزه چالش بزرگی وجود دارد که سازمان‌ها، شرکت‌ها، سیستم‌ها و داده‌ها را فرا گرفته است. این چالش ارائه راهکار مناسب و مفید و به موقع در مقابل جلوگیری و از بین رفتن و سوء استفاده داده‌ها می‌باشد. سیاست‌گذاری‌های امنیتی باعث می‌شود تا بدافزارها به درستی تشخیص داده شوند و بعد از آن مراحل برخورد با موفقیت طی شود. تصمیمات و شناسایی ضعیف می‌تواند عواقب غیرقابل جبرانی را در پیش داشته باشد. با تجزیه و تحلیل اطلاعات و داده‌ها از طریق نمودارها، عکس‌ها، متن‌ها و ... می‌توان اطلاعات با ارزشی استخراج کرد که باعث شناسایی و کشف درست بدافزار شود و زمانبندی مربوط به کشف و برخورد با بدافزار را کاهش داد. برای دستیابی به این اطلاعات باید از الگوریتم‌ها و تکنیک‌های داده‌کاوی بر روی داده‌های موجود استفاده کرد که اولین و مهمترین مرحله شناسایی فعالیت‌های بدافزار می‌باشد. یکی از مشکلات اساسی در تشخیص بدافزارها، پیچده‌تر شدن آنها توسط تکنیک‌های مبهم‌سازی است که در این صورت مقابله با بدافزارها نیازمند تشخیص و تمایز میان بدافزارها و نرم‌افزارهای قانونی است. در این راستا استفاده از هوش مصنوعی بعنوان یکی از روش‌های نوظهور و امیدوار کننده توانسته است کاربرد بسیاری جهت شناسایی و تشخیص انواع بدافزارها داشته باشد. یادگیری ماشین به عنوان یکی از شاخه‌های وسیع و پرکاربرد هوش مصنوعی، به تنظیم و اکتشاف شیوه‌ها و الگوریتم‌هایی می‌پردازد که براساس آنها، سیستم‌ها توانایی یادگیری پیدا می‌کنند و قادر خواهند انواع بدافزار را شناسایی و تشخیص دهند. لذا در این مقاله سعی نمودیم جهت تشخیص انواع بدافزارها در سیستم‌های رایانه‌ای از الگوریتم‌های چندگانه شبکه عصبی مصنوعی، ماشین بردار پشتیبان و درخت تصمیم مبتنی بر رای‌گیری حداکثری استفاده نماییم. نتایج بدست آمده حاکی از آن است که روش پیشنهادی با دقت ۹۵.۱۷ درصد برای داده‌های آموزشی و دقت ۹۳.۷۴ درصد برای داده‌های آزمایشی قادر به تشخیص انواع بدافزارها در سیستم‌های رایانه‌ای می‌باشد. بنابراین با توجه به نتایج به دست آمده می‌توان گفت روش پیشنهادی نسبت به روش‌های منفرد از عملکرد و کارایی خوبی برخوردار است.

مراجع

[1] Yong Wong, M., Landen, M., Antonakakis, M., Blough, D. M., Redmiles, E. M., & Ahamad, M. (2021, November). An inside look into the practice of malware analysis. In Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (pp. 3053-3069).

[2] Jamalpur, S. Navya, Y.S. Raja, P. Tagore, G. & Rao, G.R. (2018), «Dynamic Malware Analysis Using Cuckoo Sandbox, Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 1056-1060, Coimbatore, India.

[3] Ahmadi, M., Ulyanov, D., Semenov, S., Trofimov, M., & Giacinto, G. (2016, March). Novel feature extraction, selection and fusion for effective malware family classification. In Proceedings of the sixth ACM conference on data and application security and privacy (pp. 183-194).

[4] Dutta, N., Jadav, N., Tanwar, S., Sarma, H. K. D., Pricop, E., Dutta, N., ... & Pricop, E. (2022). Introduction to malware analysis. Cyber Security: Issues and Current Trends, 129-141.

[5] Damshenas, M., Dehghantanha, A., Choo, K. K. R., & Mahmud, R. (2015). M0droid: An android behavioral-based malware detection model. Journal of Information Privacy and Security, 11(3), 141-157.

[6] Pachhala, N., Jothilakshmi, S., & Battula, B. P. (2021, October). A comprehensive survey on identification of malware types and malware classification using machine learning techniques. In 2021 2nd international conference on smart electronics and communication (ICOSEC) (pp. 1207-1214). IEEE.

[7] Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *Ieee Access*, *8*, 75264-75278.

[8] Faruk, M. J. H., Shahriar, H., Valero, M., Barsha, F. L., Sobhan, S., Khan, M. A., ... & Wu, F. (2021, December). Malware detection and prevention using artificial intelligence techniques. In 2021 IEEE International Conference on Big Data (Big Data) (pp. 5369-5377). IEEE.

[9] Weng, C. H., Huang, T. C. K., & Han, R. P. (2016). Disease prediction with different types of neural network classifiers. Telematics and Informatics, 33(2),277-292.

[10] Peng, T., Zhou, J., Zhang, C., & Fu, W. (2017). Streamflow Forecasting Using Empirical Wavelet Transform and Artificial Neural Networks. Water, 9(6), 406.

[11] Faghihnia, E., Farizani, S. R. K. T., & Kheirabadi, M. (2021). Improved intrusion detection system based on distributed self-adaptive genetic algorithm to solve support vector machine in form of multi kernel learning with auto encoder. *Journal of Intelligent Procedures in Electrical Technology*, *12*(45), 77-93.

[12] Samantray, O. P., & Tripathy, S. N. (2020). A Knowledge-Domain Analyser for Malware Classification. In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA) (pp. 1-7). IEEE.