DOR: 20.1001.1.2538161.1404.10.4.8.0

DOI : 10.22034/csj.2025.237382

زمستان ۱۴۰۴، دوره ۱۰، شماره ۴، پیاپی ۳۹

دریافت مقاله: ۱۴۰۳/۱۰/۰۸

پذیرش مقاله: ۱۴۰۳/۱۲/۱۱

نوع مقاله: پژوهشی

تشخیص فیشینگ بر اساس الگوریتم بهینه‌سازی هریس‌هاکس و هستی‌شناسی مبتنی بر وب

سیدمحمد جوادی مقدم*

دانشکده مهندسی، دانشگاه بزرگمهر قائنات، قاین، ایران

پست الکترونیکی: smjavadim@buqaen.ac.ir

شادی دالوند

دانشکده مهندسی، دانشگاه آزاد اسلامی واحد فردوس، ایران

پست الکترونیکی: shadi_dalvand@yahoo.com

چکیده

حملات فیشینگ تله‌گذاری نوعی حمله مبتنی بر فریب است که مبتنی بر مهندسی اجتماعی یا حملات بدافزار است. اخیراً، تکنیک‌های یادگیری ماشین برای شناسایی فیشینگ نتایج امیدوارکننده‌ای را نشان داده‌اند. یک چالش برای شناسایی صفحات جعلی، انتخاب ویژگی‌های یادگیری ماشین و انتخاب بهینه ویژگی‌ها برای کاهش خطای طبقه‌بندی و زمان اجرا است. استفاده از هستی‌شناسی صفحه وب یک رویکرد عملی برای شناسایی یک حمله فیشینگ است. این مقاله یک مدل ترکیبی را ارائه می‌کند که ویژگی‌های ضروری مربوط به هستی‌شناسی صفحه وب را با استفاده از الگوریتم بهینه‌سازی هریس هاکس۱ برای یادگیری ماشین برای شناسایی حملات فیشینگ شناسایی می‌کند. روش پیشنهادی از الگوریتم هریس‌هاکس برای انتخاب ویژگی‌ها و از هستی‌شناسی صفحات وب برای کاهش خطاهای طبقه‌بندی در تشخیص حملات فیشینگ استفاده می‌کند. نتایج نشان داده‌اند که روش پیشنهادی در تشخیص حملات دقیق‌تر از روش‌های یادگیری ماشینی مانند شبکه‌عصبی مصنوعی، درخت تصمیم و جنگل تصادفی است. علاوه بر این، روش پیشنهادی دقیق‌تر از چندین روش ترکیبی انتخاب ویژگی، مانند تحلیل مولفه اصلی همراه با درخت تصمیم، جنگل تصادفی و شبکه بیزی است. نتایج ارزیابی نشان داد که روش پیشنهادی دارای دقت ۹۹/۶۸ درصد است در حالی که جذر میانگین مربعات خطا۲ ۰/۰۹۸ است.

واژه‌های کلیدی: هستی‌شناسی، تشخیص فیشینگ، الگوریتم بهینه‌سازی هریس‌هاکس، انتخاب ویژگی، صفحه وب جعلی.

مقدمه

در دنیای امروز شبکه‌های کامپیوتری نقش اساسی در ارتباطات دارند. شبکه‌های رایانه‌ای مانند اینترنت اشیاء در

* نویسندهٔ مسئول

1-Harris-Hawks Optimization(HHO)

2-RMSE

سال‌های اخیر به‌طور تصاعدی رشد کرده‌اند [۱]. در اکثر شبکه‌های کامپیوتری، امنیت یک موضوع حیاتی است زیرا اگر امنیت شبکه حفظ نشود، عملکرد شبکه کاهش می‌یابد و اطلاعات افراد به سرقت می‌رود [۲]. حملات مختلفی مانند DDoS[۳][۳]، حملات متقابل[۴] [٤]، سرقت‌های سایبری[۵] [۵]، و بدافزار[۶] [٦] می‌تواند شبکه‌های نسل جدید را تحت تأثیر قرار دهد.

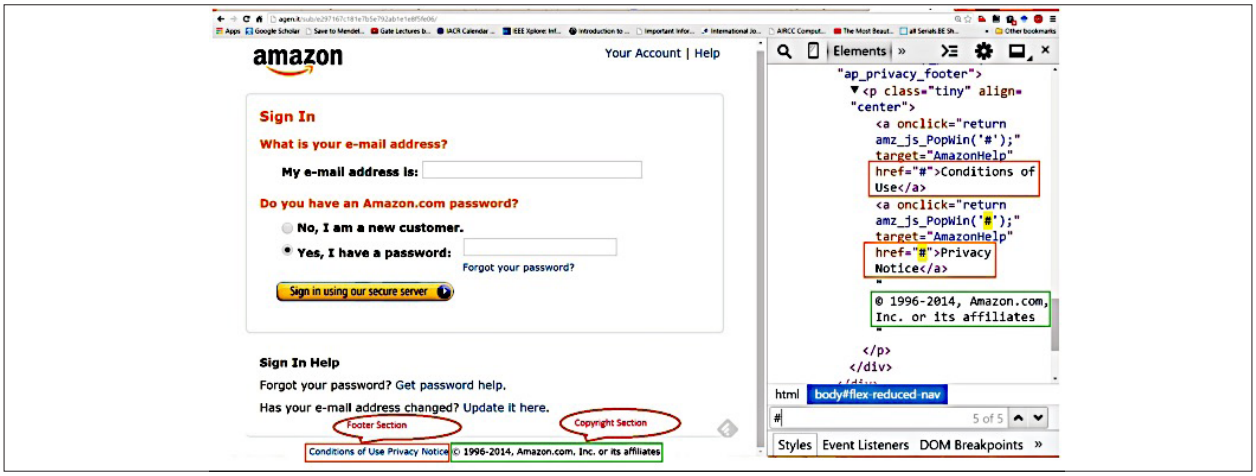
حملات فیشینگ [۷] کاربران و مردم را در اینترنت تهدید می‌کند. گاهی اوقات، یک پیوند[۷] برای کاربران اینترنت اشیا ارسال می‌شود و آنها را به یک صفحه جعلی هدایت می کند. یک چالش در شناسایی حملات فیشینگ این است که بسیاری از اتصالات یا صفحات وب نیاز به تجزیه‌وتحلیل دارند. بررسی این حجم زیاد از داده‌های مرتبط با وب زمان زیادی می‌برد.

سه روش اصلی برای تشخیص صفحات جعلی از واقعی در اینترنت [۸] وجود دارد. استفاده از لیست سیاه [۹] اولین روش برای تشخیص فیشینگ است. در روش لیست سیاه، نشانی وبگاههای جعلی نگهداری می‌شود. تمام محتوای لیست سیاه برای کشف الگوی صفحات جعلی برای شناسایی حملات فیشینگ بررسی می‌شود. روش‌های لیست سیاه فاقد هوش هستند و به حافظه زیادی نیاز دارند. در تکنیک دوم، به نام روش‌های اکتشافی [۱۰]، از شواهدی مانند طول نشانی یا تعداد زیردامنه‌ها برای شناسایی حملات فیشینگ استفاده می‌شود، اما این روش‌ها نیز تقریبی هستند. برخلاف دو رویکرد قبلی در یادگیری ماشینی [۱۱] و رویکرد یادگیری عمیق [۱۲]، امکان کشف الگوی صفحات جعلی با آموزش الگوریتم یادگیری ماشین وجود دارد. به همین دلیل، این روش‌ها در اکثر مطالعات برای شناسایی صفحات جعلی، مانند شبکه‌های عصبی مصنوعی [۱۳]، درخت‌های تصمیم [۱٤]، و جنگل‌های تصادفی [۱۵] استفاده می‌شوند. یکی از بحرانی‌ترین مشکلات در شناسایی وبگاه‌های جعلی در اینترنت با روش‌های یادگیری ماشینی، تعداد زیاد ویژگی‌ها است. ویژگی‌های بهینه برای تشخیص صفحات جعلی باید در یادگیری ماشین در نظر گرفته شود.

ناگونوا و همکاران [۱٦] از روشی هوشمند مبتنی بر ویژگی‌های جدید برای شناسایی حملات فیشینگ و وبگاه‌ها استفاده کردند. حکیم و همکاران [۱۷] مکانیزمی را برای شناسایی و ارزیابی حملات فیشینگ با استفاده از اندازه‌گیری ایمیل ارائه کردند که روشی مناسب برای مطالعه علوم رفتاری افراد در تشخیص فیشینگ می‌باشد. برخی از محققان [۱۸, ۱۹] رویکردی مبتنی بر الگوریتم‌های ژنتیک برای طبقه‌بندی و وبگاه‌های جعلی و واقعی ارائه کردند. زو و همکاران [۲۰] مدلی را با استفاده از درخت تصمیم با ویژگی‌های بهینه ارائه کردند. بازکر [۲۱] یک روش تشخیص پردازش لوگو را با استفاده از یک هیستوگرام مبتنی بر گرادیان برای شناسایی وب‌گاه فیشینگ و نام تجاری جعلی پیشنهاد کرد. آزمایش‌ها نشان می‌دهند که در بهترین پیکربندی، برنامه این روش پیشنهادی یا 'LogoSENSE' می‌تواند به دقت ۹۳/۵۰ درصد دست یابد.

الجوفی[۲۲] رویکردی را با استفاده از URL[۸] و HTML[۹] وبگاه‌ها ارائه کرد. روش پیشنهادی به دقت ۹٦/۷٦ درصد در تشخیص صفحات فیشینگ رسیده است. ژو [۱٤] از یک تکنیک بهینه‌سازی تکاملی چند هدفه اصلاح شده همراه با روش جنگل تصادفی استفاده کرد. مینوچا [۲۳] یک بهینه ساز تعادل اصلاح شده دودویی ارائه کرد که ویژگی‌های اساسی صفحات وب فعلی را کشف می‌کند و از طبقه‌بندی‌کننده نزدیکترین همسایه برای طبقه‌بندی صفحات جعلی و اصلی استفاده می‌کند. چند محقق رائو و پایس [۲٤]، یریما و الزیلایی [۲۵] و ماشاله [۲٦] روش‌هایی را برای نشانی‌های فیشینگ با استفاده از شبکه عصبی

3-Distributed Denial of Service
4-Crossfire
5-Cyber thefts
6-Malware
7-Link
8-Uniform Resource Locator
9-Hypertext Markup Language



شکل ۱: لپیوندهای خالی در کد منبع سایت نشانه فیشینگ است.

هم‌آمیختی[10] ارائه کردند. شبکه‌های پیشنهادی به طور بهینه در دستگاه‌های تلفن همراه با عملکرد مناسب مورد استفاده قرار می‌گیرند.

تمام روش‌های ذکر شده سعی در افزایش دقت و کاهش پیچیدگی زمانی تشخیص فیشینگ داشتند. چالش شبکه عصبی این است که اگر با تمام ویژگی‌های صفحات جعلی آشنا شود، به زمان زیادی برای آموزش نیاز دارد و خطای یادگیری افزایش می‌یابد. با انتخاب ویژگی‌ها می‌توان یادگیری را بر روی ویژگی‌های ضروری متمرکز کرد و خطای تشخیص حملات و شناسایی صفحات جعلی و اصلی را کاهش داد. شناسایی ویژگی‌های مرتبط با هستی‌شناسی صفحات وب و پیوندهای وب برای یادگیری شبکه‌های عصبی به منظور کاهش خطاهای تشخیص حمله یکی از اهداف ضروری این تحقیق است.

این مقاله یک الگوریتم بهبود یافته بهینه‌سازی هریس هاکس [۲۷] را معرفی می‌کند تا بهترین ویژگی‌ها را برای یک شبکه عصبی مصنوعی برای انتخاب ویژگی در تشخیص صفحات جعلی در اینترنت ارائه کند. ادامه این مقاله به شرح زیر است: بخش بعدی صفحات جعلی در اینترنت را شرح می‌دهد. بخش سه روش پیشنهادی، بخش چهار نتایج و بحث و در نهایت، بخش پنج نتیجه‌گیری را ارائه می‌دهد.

10-Convolution

۲- صفحه جعلی

وبگاههای جعلی گروهی از وبگاهها هستند که شبیه وبگاههای قانونی برای سرقت اطلاعات کاربران هستند. ظاهر صفحات جعلی شباهت زیادی به صفحات قانونی دارد و این صفحات دارای امکاناتی مشابه وبگاه‌های واقعی هستند، اما دارای ویژگی‌هایی هستند که می‌توان تا حدودی جعلی بودن آنها را تشخیص داد. شکل ۱ یک صفحه فیشینگ و جعلی را نشان می‌دهد که می‌توان از آن برای تعیین جعلی بودن آن استفاده کرد. در این مثال مشاهده می‌شود که تعداد پیوندهای خالی در صفحات جعلی بیش از حد است و عاملی حیاتی در شناسایی فیشینگ [۲۸] است.

مثال دیگر جعل پیوندهای اینترنتی است. پیوندهای اینترنتی جعلی به نشانی دامنه یک وبگاه قانونی، زیردامنه نشانی جعلی در نظر گرفته می‌شود [۲۹]. در برخی موارد، رخنه‌گر با کمک املای غلط و جابه‌جایی کلمات یک نشانی جعلی ایجاد می‌کند تا نشانی جعلی کمی با نشانی وبگاه قانونی متفاوت باشد. گاهی اوقات، رخنه‌گر نشانی قانونی را در قسمت اول نشانی قرار می‌دهد. سپس نشانی جعلی را در قسمت دوم یا بعد از نویسهٔ @ قرار می‌دهد. اکثر مرورگرها نشانی قبلی این نویسه را نادیده می‌گیرند و از نشانی بعدی استفاده می‌کنند. گاهی اوقات، رخنه‌گر از نشانی‌های IP، اعداد مبنای شانزده و اعشاری یا حتی ترکیبی برای گمراه کردن کاربران استفاده می‌کند. یک

تله‌گذار[11] می‌تواند نشـــانی‌های جعلی را در یک ایمیل قرار داده و برای کاربران ارسال کند. کاربران با کلیک بر روی این پیوند جعلی وارد وبگاه فیشینگ می‌شـــوند و اطلاعات آنها به ســـرقت می‌رود. در موارد دیگـــر، کاربر وارد یک وبگاه فیشینگ شده و با استفاده از کد جاوا اسکریپت بدون کلیک بر روی پیوند جعلی وارد وبگاه جعلی می‌شود [۳۰].

۳- روش پیشنهادی

یـــک چالش اســـاسی در تشـــخیص وبگاه‌های جعلی، مســـئله انتخاب ویژگی اســـت که بر دقت طبقه‌بندی تأثیر می‌گذارد. روش پیشنهادی از الگوریتم هریس‌هاکس برای انتخاب ویژگی‌ها و از هستی‌شـــناسی صفحات وب برای کاهش خطاهای طبقه‌بندی در تشـــخیص حملات فیشینگ اســـتفاده می‌کند. شـــکل ۲ روندنمای روش پیشنهادی را نشان می‌دهد.

یـــک بردار ویژگی را می‌توان بـــا میانگین خطا و تعداد ویژگی‌هـــای انتخاب شـــده در هر بـــردار ارزیـــابی کرد. خروجی هر تکرار الگوریتم هریس‌هاکس یک بردار ویژگی بهینه در نظر گرفته می‌شود. ســـایر بردارهای ویژگی در اطراف راه‌حل بهینه جستجو می‌کنند تا بهینه‌ترین راه‌حل را پیدا کنند. مراحل زیر برای بردارهای ویژگی اعمال می‌شود تا خطای تشخیص فیشینگ آنها کاهش یابد تا آنها به روز شوند:

- بردارهای ویژگی را به صورت تصادفی به روز کنید
- بردارهای ویژگی را با محدودیت نرم[12] به روز کنید
- بردارهای ویژگی را با محدودیت سخت[13] به روز کنید
- بردارهای ویژگی را با محدودیت سریع[14] به روز کنید

بردار ویژگی بهینـــه را می‌توان یک بردار از نوع طعمه در نظـــر گرفت که ســـایر بردارهای ویژگی بـــه دنبال به روزرسانی آن هستند.

۳-۱- هریس‌هاکس

در روش پیشـــنهادی، مراحل زیر بـــرای انتخاب بردار ویژگی بهینه با یک خطای جزئی برای تشـــخیص فیشینگ تکرار می‌شوند:

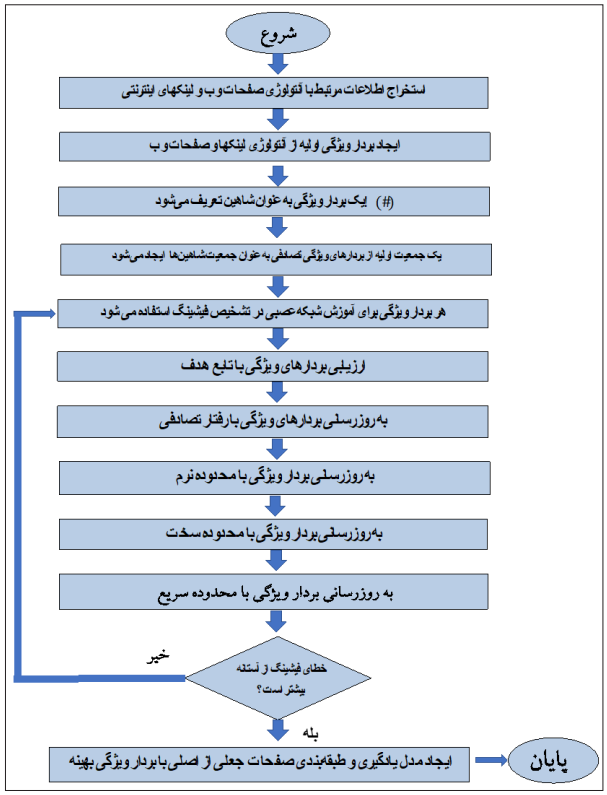
- بردار ویژگـــی با مقادیر صفر و یک تعریف می‌شـــود. مقدار صفر نشـــان می‌دهد که هیچ ویژگی انتخاب نشـــده است و مقدار یک نشان‌دهنده ویژگی انتخاب شده است.
- تعدادی از این بردارهای ویژگی به صورت تصادفی به عنوان جمعیتی از الگوریتم هریس‌هاکس تولید می‌شـــوند. هـــر بردار ویژگی می‌تواند انـــدازه مجموعه داده را کاهش دهد. سپس، مجموعه داده‌های کاهش‌یافته به‌عنوان ورودی یک روش طبقه‌بندی مانند شبکه عصبی مصنوعی در نظر گرفته می‌شـــوند. این ویژگی‌های انتخاب شـــده روش‌های یادگیری ماشینی را آموزش می‌دهند.
- بـــردار با حداقل خطا و تعداد ویژگی‌های انتخاب شـــده بردار ویژگی بهینه فرض می‌شود.
- بردار ویژگی بهینـــه در الگوریتم هریس‌هاکس موقعیت طعمه در نظر گرفته می‌شود. ســـایر بردارها یا شاهین‌ها نیز در اطراف آن جســـتجو می‌شوند. بردارهای ویژگی با اســـتفاده از بردار ویژگی یا موقعیت طعمـــه بهینه به‌روز می‌شوند.
- بردارهای ویژگی به‌روز شـــده شبکه عصبی مصنوعی چند لایه را در هر مرحلـــه آموزش می‌دهند. بردار ویژگی بهینه بردار ویژگی است که دارای حداقل خطا است.

مراحل موردنظر تکرار می‌شوند و در نهایت در آخرین تکرار، بردار ویژگی بهینه انتخاب شده و بر روی مجموعه داده‌های فیشینگ اعمال می‌شود و سپس به عنوان ورودی طبقه‌بندی استفاده می‌شود.

۳-۲- تابع هدف

یک بردار ویژگی دارای n جزء یا ســـلول اســـت و هر ســـلول یک ویژگی اســـت که می‌تواند صفر یا یک باشـــد. ســـپس با اســـتفاده از الگوریتم هریس‌هاکس که دارای m

11-Phisher
12-Smooth siege
13-Hard siege
14-Fast siege



شکل ۲: روندنما روش پیشنهادی برای تشخیص حملات فیشینگ با انتخاب ویژگی‌ها و هستی‌شناسی

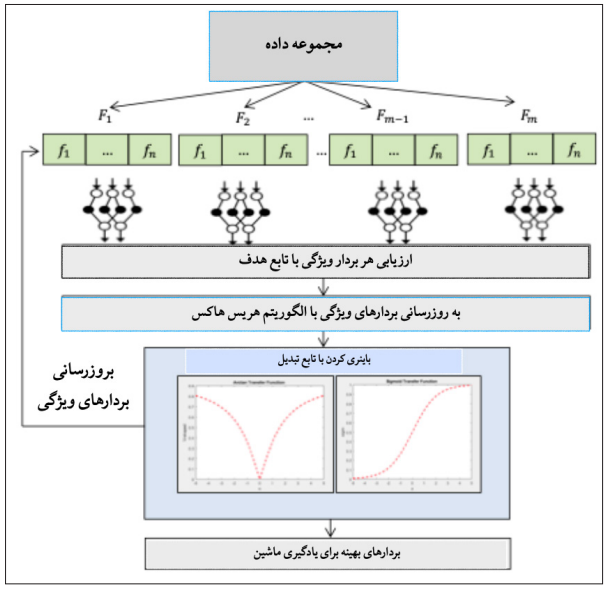
ویژگی است، ویژگی بهینه انتخاب می‌شود. معمولاً n>m است و ارزیابی میانگین خطای تشخیص فیشینگ و تعداد ویژگی‌های انتخاب شده برای ارزیابی هر بردار ویژگی در هر تکرار ضروری است. یک شاهین راه‌حلی برای مشکل تعریف می‌کند. علاوه بر این، موقعیت طعمه راه‌حل بهینه فعلی است. شاهین‌ها سعی می‌کنند به سمت طعمه پرواز کنند. در الگوریتم هریس‌هاکس، شاهین‌ها طعمه را در فضای مشکل جستجو می‌کنند و سپس به آن حمله می‌کنند. در روش پیشنهادی، یک بردار ویژگی یک شاهین در نظر گرفته شده و برای آموزش شبکه عصبی مصنوعی استفاده می‌شود. یک بردار ویژگی در تکرار فعلی یا t برابر با X (t) فرض می‌شود و مقدار به‌روز شده آن در تکرار جدید X (t + 1) است. معادله (۱) تابع هدف برای ارزیابی یک بردار ویژگی است:

$$Cost(X(t)) = \frac{1}{2}\times\frac{1}{n}\sum_{i=1}^{n}|\overline{Y}_i\text{-}Y_i| + \frac{1}{2}\times\frac{f}{F} \quad (۱)$$

در این رابطه، مقدار تابع انتخاب ویژگی یک بردار ویژگی مانند $X(t)$ برابر $Cost(X(t))$ است و در اینجا مقدار خطای طبقه‌بندی برای تشخیص فیشینگ برابر $\frac{1}{n}\sum_{i=1}^{n}|\overline{Y}_i\text{-}Y_i|$ است که در آن n تعداد نمونه‌ها است و از طرفی هم $\overline{Y}_i$ مقدار پیش‌بینی یک نمونه و Y_i مقدار واقعی یک نمونه است که می‌تواند فیشینگ یا وبگاه عادی و قانونی باشد. در این رابطه، f تعداد ویژگی انتخاب شده برای ویژگی مانند $X(t)$ است و F تعداد کل ویژگی‌های ممکن برای تشخیص فیشینگ است. هر بردار ویژگی که این تابع هدف را کمینه‌تر نماید بردار ویژگی بهینه است که با $X_{rabbit}(t)$ نمایش داده می‌شود و در واقع موقعیت طعمه است.

بردار ویژگی دارای ۳۰ ویژگی است که برخی به هستی‌شناسی صفحات وب و پیوندهای وب و برخی مربوط به ویژگی‌های مرتبط با موتورهای جستجو هستند. یک بردار ویژگی می‌تواند دارای مقادیر ۱+، ۱− و صفر باشد. برخی از ویژگی‌ها، مانند طول عمر دامنه، می‌توانند کوتاه، متوسط یا طولانی باشند. سه مقدار به ترتیب با ۱−، ۰ و ۱+ نشان داده می‌شوند. برخی از ویژگی‌ها دارای دو مقدار هستند، مانند داشتن نشانی IP. اگر مقدار حوزهٔ IP به صورت صفر نمایش داده شود، نشانی IP ندارد. اگر یک باشد، نشانی IP دارد. مطابق شکل ۳، فرایند زیر برای ایجاد بردارهای ویژگی اولیه بر اساس هستی‌شناسی صفحات و پیوندها انجام می‌شود:

- ابتدا چندین بردار ویژگی تصادفی از ویژگی‌های مجموعه داده ایجاد می‌شود.
- هر بردار ویژگی عضوی از الگوریتم هریس‌هاکس است و مشخص می‌کند که از کدام ستون‌های داده استفاده می‌شود.
- در هر مرحله، بردارهای ویژگی با شبکه عصبی ارزیابی می‌شوند.
- بردارهای ویژگی با الگوریتم هریس‌هاکس به روز می‌شوند.
- با توابع تبدیل، بردارهای پیوسته به مقادیر دودویی تبدیل می‌شوند.



شکل ۳: ایجاد بردارهای ویژگی اولیه بر اساس هستی شناسی صفحات و پیوندها

۴- نتایج

۴-۱- مجموعه داده

برای ارزیابی روش پیشنهادی از مجموعه داده یادگیری ماشین UCI استفاده شده است [۳۱]. این مجموعه داده دارای بیش از ۱۱۰۰۰ رکورد با ۳۰ ویژگی ورودی است. علاوه بر این، ویژگی سی و یکم خروجی است که می‌تواند نشان‌دهنده فیشینگ یا عدم فیشینگ یک وبگاه باشد. این مجموعه داده دارای چهار دسته ویژگی کلی زیراست:

- ویژگی‌های مربوط به پیوند یا نشانی
- ویژگی‌های مربوط به کد منبع وبگاه
- ویژگی‌های مرتبط با دامنه وبگاه
- ویژگی‌های مختلف، مانند استفاده از برچسب‌های منحصر به فرد و یا ویژگی‌های مرتبط با موتورهای جستجو

- چند ویژگی صفحات فیشینگ مربوط به هستی‌شناسی صفحات وب و پیوندهای اینترنتی به شرح زیر است:

- استفاده از نشانی اینترنتی یا IP یکی از ویژگی‌های اساسی صفحات جعلی است که نمونه‌ای از آن را می‌توان در http://125.98.3.123/fake.html مشاهده کرد.
- استفاده از یک URL طولانی برای پنهان کردن بخش‌های مشکوک در URL وبگاه یکی از روش‌هایی است که توسط رخنه‌گران برای جلوگیری از سرقت اطلاعات افراد استفاده می‌شود و قانون آن به شرح زیر است:

$$\begin{cases} URL\ length < 54 \rightarrow feature = \text{Legitimate} \\ else\ if\ URL\ length \geq 54\ and\ \leq 75 \rightarrow feature = Suspicious \\ otherwise \rightarrow feature = \text{Phishing} \end{cases}$$

- استفاده از URLهای مختصر می‌تواند کاربران را فریب دهد تا درگیر سرقت برخط شوند. به عنوان مثال، URL وبگاهی مانند http://portal.hud.ac.uk به طور خلاصه bit.ly/19DXSk4 نشان داده می‌شود.
- وجود نویسهٔ @ در نشانی، نشانه فیشینگ در وبگاه‌ها و نشانی‌های جعلی است و با کمک این نویسه، رخنه‌گر اطلاعات ضروری را برای افراد ایمیل می‌کند.
- برای فریب کاربران و اتصال دامنه‌های جعلی به دامنه‌های مهم، مانند http://www.Confirme- paypal.com، زیردامنه‌هایی با نویسهٔ "-" ایجاد می‌کند.
- بسیاری از زیردامنه‌ها در یک نشانی وبگاه مانند http://www.hud.ac.uk/students/ نشانه حملات فیشینگ است.
- اگر زمان مورد نیاز برای ارسال گواهی رمزگذاری پروتکل Htpps کمتر از زمان مشخص شده باشد، نمی‌توان به وبگاه اعتماد کرد.
- اگر مدت زمان ثبت نشانی دامنه طبق موارد زیر کمتر از یک سال است.

قانون زیر نشان می‌دهد که وبگاه اعتبار کمی دارد:

$$\begin{cases} \text{Domains Expires on} \leq 1\ \text{years} \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{cases}$$

- عدم وجود تصاویر کوچک در قسمت URL نشان می‌دهد که وبگاه از اعتبار کمی برخوردار است.
- عدم وجود نماد فاویکون[۱۵] در URL نشان می‌دهد که وبگاه اعتبار کمی دارد.

15-Favicon icon

● استفاده از شماره درگاه‌های نامتعارف در نشانی وبگاه نشان می‌دهد که نشانی مشکوک و احتمالاً فیشینگ است.

● از عبارت HTTPS به عنوان یک زیردامنه استفاده کنید تا کاربران فریب بخورند و فکر کنند وبگاه موردنظر رمزگذاری شده است، مانند http://https-www- paypal.com/

● ارجاع وبگاه‌های دیگر بـــه وبگاه‌های دیگر در صورتی که مطابقت با قانون زیر، کمتر از ۲۲ است، وبگاه موردنظر قانونی است و اگر بین ۲۲ و ۶۱ باشد مشکوک است و اگر بیشتر از ۶۱ باشد، وبگاه موردنظر فیشینگ است:

$$\begin{cases} \%\ \text{of Request URL} < 22\% \rightarrow \text{Legitimate} \\ \%\text{of Request URL} \geq 22\% \text{ and } 61\% \rightarrow \text{Suspicious} \\ \text{Otherwise} \rightarrow \text{feature} = \text{Phishing} \end{cases}$$

● اگر تعداد زیـــادی URL و پیوند خالـــی در وبگاه وجود دارد، وبگاه جعلی اســـت. با توجه به قانون زیر، اگر تعداد پیوندهای خالی کمتر از ۳۰ درصد باشـــد وبگاه موردنظر نرمال اســـت واگر بین ۳۱ تا ۶۷ باشد مشکوک است و اگر بیشتر از این باشد فیشینگ است.

$$\begin{cases} \%\ \text{of URL Of Anchor} < 31\% \rightarrow \textit{Legitimate} \\ \%\ \text{of URL Of Anchor} \geq 31\% \text{ And} \leq 67\% \rightarrow \text{Suspicious} \\ \text{Otherwise} \rightarrow \text{Phishing} \end{cases}$$

● استفاده از برچسب‌های منحصر به فرد که رایج نیستند می‌تواند فیشینگ را نشان می‌دهد.

● اگر نشانی وبگاه از دامنه‌های غیرمعمول استفاده می‌کند کـــه این کار را انجام می‌دهند دامنه را با دســـتور WHOIS فراخوانی کنید، وبگاه به احتمال زیاد فیشینگ است.

● در وبگاه‌هـــای فیشـــینگ میـــزان انتقال بـــه وبگاه‌ها و دامنه‌های دیگر خارج از دامنه فعلی بســـیار است و قانون به شرح زیر است:

$$\begin{cases} \text{ofRedirect Page} \leq 1 \rightarrow \text{Legitimate} \\ \text{of Redirect Page} \geq 2 \text{ And} < 4 \rightarrow \text{Suspicious} \\ \text{Otherwise} \rightarrow \text{Phishing} \end{cases}$$

● جعل نشانی وبگاه یا پیوند در نوار وضعیت مرورگر با جاوا اسکریپت نیز یکی از روش‌های فیشینگ است.

● غیرفعال کردن کلیک راســـت توسط جاوا اسکریپت یک نشانه اصلی فیشینگ است.

● اســـتفاده از پنجره‌های تبلیغاتی با ورود اطلاعات کاربر یک روش رایج فیشینگ است.

● اســـتفاده از محتوای وبگاه توسط برچسب‌های Frame یک روش استاندارد فیشینگ است.

● عمر دامنه در وبگاه‌های فیشـــینگ اغلب کوتاه اســـت و به عنوان یک قاعده، اگر کمتر از شـــش ماه باشند، احتمال فیشینگ بیشتر است:

$$\begin{cases} \text{Age Of Domain} \geq 6\ \text{months} \rightarrow \text{Legitimate} \\ \text{Otherwise} \rightarrow \text{Phishing} \end{cases}$$

● عـــدم وجود رکورد دامنه در یک وبگاه به کمک WHOIS بر اساس قانون ذیل می‌تواند نشان‌دهنده فیشینگ باشد:

$$\begin{cases} \text{no DNS Record For The Domain} \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{cases}$$

● وبگاه‌های فیشـــینگ و تقلبی ترافیک کمتری نســـبت به وبگاه‌های اصلـــی و قانونی دارند. میـــزان ترافیک صفحه وب در موتورهای جســـتجو مانند الکســـا[16] یک شاخص ضروری در تشخیص فیشینگ است:

$$\begin{cases} \text{Website Rank} < 100{,}000 \rightarrow \text{Legitimate} \\ \text{Website Rank} > 100{,}000 \rightarrow \text{Suspicious} \\ \text{Otherwise} \rightarrow \text{Phish} \end{cases}$$

● رتبه صفحه در گوگل یک شاخص مهم است که می‌تواند به تعیین این که آیا وبگاه فیشـــینگ اســـت کمک کند که بر اساس قانون زیر تعیین می‌شود:

$$\begin{cases} \text{PageRank} < 0.2 \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{cases}$$

● گوگل معمولاً صفحات فیشـــینگ را فهرست نمی‌کند، در حالی که صفحات قانونی در اینترنت ایندکس می‌شوند.

● داشتن تعداد زیادی پیوند به یک صفحه می‌تواند اعتبار آن را نشان دهد. طبق قانون، وبگاه‌های فیشینگ به دیگران پیوند داده نمی‌شوند.

طبـــق این قانون اگر وب‌گاهی بیش از دو پیوند دریافت کند، احتمالا قانونی است:

$$\begin{cases} \text{Of Link Pointing to The Webpage} = 0 \rightarrow \text{Phishing} \\ \text{Of Link Pointing to The Webpage} > 0 \text{ and} \leq 2 \rightarrow \text{Suspicious} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{cases}$$

16-Alexa

اطلاعات وبگاه‌های امنیتی مانند PhishTank در محیط اینترنت سالانه و ماهانه می‌تواند برای تجزیه‌وتحلیل وبگاه‌های مختلف حیاتی باشد.

۴-۲- معیارهای ارزیابی

ارزیابی خطا و شاخص‌های طبقه‌بندی برای ارزیابی هر روش یادگیری ماشین در تشخیص حملات فیشینگ در نظر گرفته شد. خطای جذر میانگین مربعات یا RMSE برای شناسایی فیشینگ و تجزیه‌وتحلیل هر یک از روش‌های یادگیری ماشین مطابق با معادله ۲ استفاده شد:

$$rmse = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\overline{Y}_i\text{-}Y_i)^2} \qquad (۲)$$

در این رابطه، $\overline{Y}_i$ مقدار واقعی یک نمونه و Y_i مقدار تخمینی یک نمونه است و از طرفی m تعداد نمونه‌هایی است که مورد ارزیابی قرار گرفته می‌شود. علاوه بر شاخص متوسط خطا می‌توان از شاخص‌های طبقه‌بندی نیز برای ارزیابی هر یک از این روش‌ها استفاده نمود که مهمترین آنها شاخص صحت[17]است که در رابطه ۳ برای تشخیص صفحات جعلی از قانونی فرموله شده است:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (۳)$$

برای محاسبه شاخص صحت نیاز است که تعداد نمونه صحیح مثبت[18](TP)، تعداد غلط منفی[19] (FN)، تعداد غلط مثبت[20] (FP) و تعداد صحیح منفی[21] (TN)، محاسبه شده و برای محاسبه شاخص‌های دقت، صحت و حساسیت مورد استفاده قرار گرفته شود و این شاخص‌ها را به طور معمول بر حسب درصد بیان نموده و مقدار بیشتر نشان‌دهنده طبقه‌بندی بهینه‌تر است.

هر کدام از شاخص‌های دقت، صحت و حساسیت بیان‌کننده یک مفهوم است که در ذیل به آنها اشاره شده است

- دقت: توانایی تشخیص صفحات جعلی و قانونی است اما مشخص نمی‌شود که کدام یک با چه دقت شناسایی شده است.
- حساسیت: توانایی تشخیص صفحات جعلی یا فیشینگ را از صفحات قانونی و اصلی نشان می‌دهد.
- صحت: نشان می‌دهد که وبگاه‌های که جعلی در نظر گرفته شده‌اند با چه دقتی از نوع وبگاه جعلی می‌باشند.

هر کدام از نمونه‌های TN،FP،TP و FN برای تشخیص حملات فیشینگ و صفحات جعلی دارای معنای مشخص به شرح ذیل می‌باشند:

- TP: یک وبگاه جعلی که روش پیشنهادی آن را جعلی شناسایی نموده است.
- FP: یک وبگاه اصلی بوده و روش پیشنهادی آن را به غلط جعلی شناسایی نموده است.
- TN: یک وبگاه اصلی بوده و روش پیشنهادی آن را به درستی اصلی شناسایی نموده است.
- FN: یک وبگاه جعلی بوده و روش پیشنهادی آن را اصلی شناسایی نموده است.

۴-۳- محیط ارزیابی

روش پیشنهادی در محیط برنامه‌نویسی متلب پیاده‌سازی شد. پارامترها به صورت زیر تنظیم شدند:

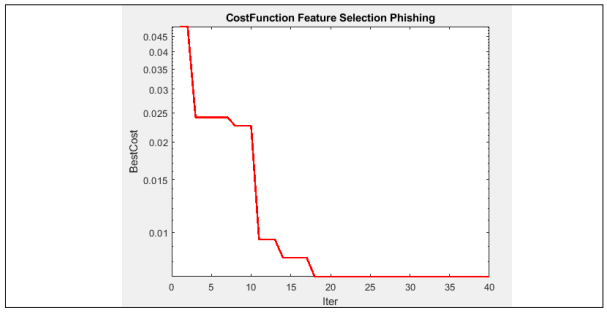
- جمعیت الگوریتم هریس هاکس ۱۰
- تعداد تکرار ٤٠
- ۷۰ درصد داده‌ها مجموعه آموزشی در نظر گرفته شد و ۳۰ درصد مجموعه تست بود.

۴- ۴- نتایج ارزیابی

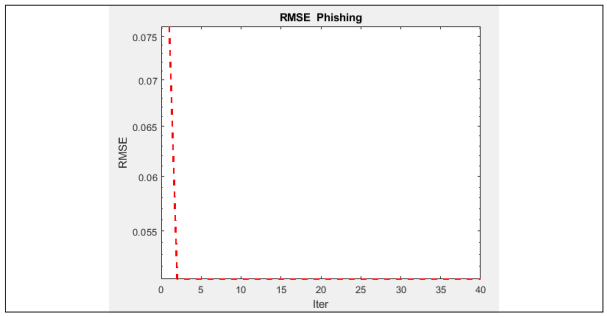
شکل‌های ٤ تا ٦ خروجی روش پیشنهادی را برای تشخیص صفحات جعلی از اصلی از نظر مقدار تابع انتخاب ویژگی، RMSE و صحت نشان می‌دهند.

خطای RMSE روش پیشنهادی در تشخیص صفحات فیشینگ با روش‌های شبکه‌عصبی مصنوعی چندلایه

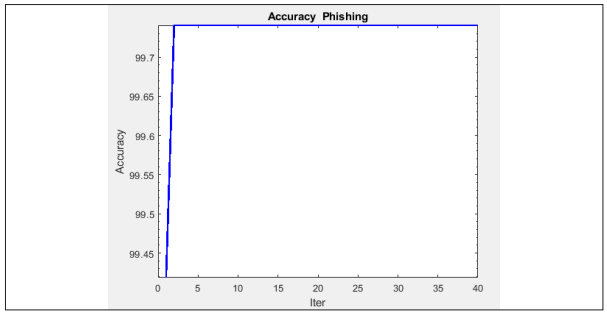
17-Accuracy
18-True Positive(TP)
19-False Negative (FN)
20-False Positive(FP)
21-True Negative (TN)



شکل ۴: تحلیل خروجی تابع انتخاب ویژگی در تشخیص صفحات جعلی از اصلی



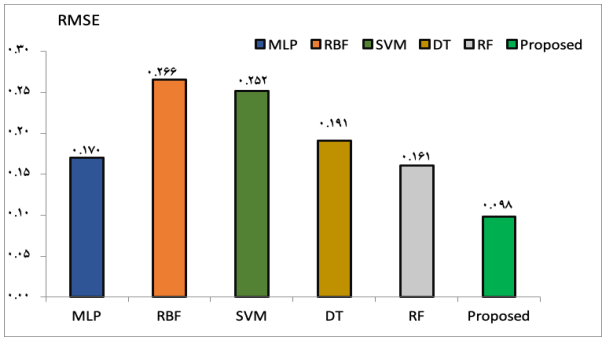
شکل ۵: تحلیل خروجی خطای RMSE تشخیص صفحات جعلی از اصلی



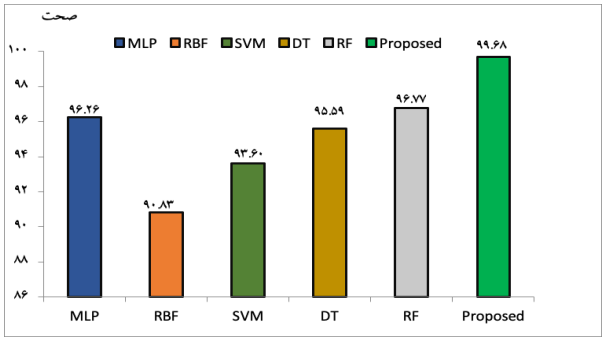
شکل ۶: تحلیل خروجی صحت تشخیص صفحات جعلی از اصلی

(MLP)، شبکه عصبی مصنوعی بازگشتی (RBF)، ماشین بردار پشتیبان (SVM)، درخت تصمیم (DT) و جنگل تصادفی (RF) مطابق نمودار (شکل ۷) مقایسه شد. آزمایش‌ها نشان داد که خطای روش پیشنهادی در تشخیص حملات فیشینگ ۰/۰۹۸ است. خطاهای شبکه‌های عصبی مصنوعی چندلایه، شبکه‌های عصبی مصنوعی بازگشتی، ماشین‌های بردار پشتیبان و جنگل‌های تصادفی برای تشخیص صفحات جعلی به ترتیب ۰/۱۷۰، ۰/۲۲۶، ۰/۲۵۲، ۰/۱۹۱ و ۰/۱۶۱ بود.

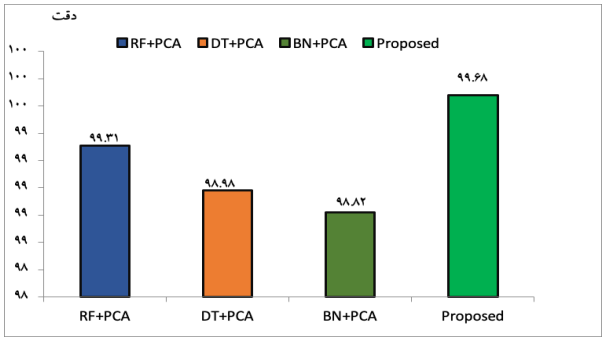
صحت روش پیشنهادی حدود ۹۹/۶۸ درصد بود و با روش‌های یادگیری پایه مقایسه شد، همان‌طور که در شکل ۸ نشان داده شده است. براساس نمودار صحت، صحت



شکل ۷: مقایسه مربعات خطا در روش پیشنهادی و چندین روش یادگیری ماشین



شکل ۸: مقایسه صحت در روش پیشنهادی و چندین روش یادگیری ماشین



شکل ۹: مقایسه صحت تشخیص صفحات جعلی در روش پیشنهادی و روش‌های یادگیری ماشین با انتخاب ویژگی PCA.

شبکه‌های عصبی مصنوعی چند لایه، شبکه‌های عصبی مصنوعی بازگشتی، ماشین‌های بردار پشتیبان و جنگل تصادفی برای تشخیص فیشینگ به ترتیب ۹۶/۲۶ درصد، ۹۰/۸۳ درصد، ۹۳/۶۰ درصد، ۹۵/۵۹ درصد و ۹۶/۷۷ درصد است.

شکل ۹ روش پیشنهادی را با روش‌های یادگیری ماشین جنگل تصادفی، درخت تصمیم و شبکه بیزی (BN) مقایسه می‌کند که از مرحلهٔ انتخاب ویژگی براساس کاهش تحلیل مؤلفه اصلی (PCA) در شناسایی حملات فیشینگ استفاده می‌کند.

۴-۵- بحث

شکل ٤ نشان می‌دهد که مقدار تابع هدف برای انتخاب ویـژگی از حـدود ۰/۰٤٥ به حـدود ۰/۰۰۹٥ کاهش یافته است. از آنجا که تابع هدف انتخاب ویژگی از دو عامل خطا و میانگین تعداد ویژگی‌های انتخاب شـده ایجاد می‌شود، نتایج زیر به دست می‌آید:

- کاهش تابع هدف نشـان می‌دهد که خطای تشخیص با توجه به تکرار الگوریتم در حال کاهش است.
- تعـداد ویژگی‌هایـی که صفحات جعلـی را از صفحات قانونی متمایز می‌کند نیز رو به کاهش است.

الگوریتـم هریس‌هاکـس روند امیدوارکننـده و بهبود مستمر را نشـان می‌دهد. با هر بار تکرار، الگوریتم تلاش می‌کنـد تا بردار ویـژگی بهینه را انتخاب کنـد، در نتیجه خطا در تشـخیص صفحـات جعلی از صفحـات قانونی کاهش می‌یابد. نتایج نشان‌دهنده کاهش قابل توجه خطای تشخیص صفحات جعلی اسـت، از حدود ۰/۰۷٥ در اولین تکـرار به تقریبـا ۰/۰٥۰ در تکرارهای بعـدی. این کاهش مداوم در خطـا گواهی بر پتانسیل الگوریتم هریس هاکس در انتخاب بردار ویژگی است.

صحت روش پیشـنهادی از نظر تکرار نیز برای تحلیل آن ضروری است. مشـاهده می‌شود که الگوریتم هریس هاکس به طور مداوم سـعی در افزایش این صحت داشته است. افزایش صحت در تکرارهای اول حدود ۹۷/٦۸ بود. با این حـال، در تکرارهای نهایی، تقریبـاً ۹۹/۷٦ بود. این افزایش صحت نیز نتیجه انتخاب ویژگی توسـط الگوریتم هریس هاکس اسـت. شاخص‌های خطا و صحت به اندازه جمعیـت در الگوریتم هریس هاکس حسـاس هسـتند. با افزایش جمعیت، شـانس یافتن بردار ویژگی بهینه افزایش می‌یابد. خطای تشـخیص فیشینگ کاهش می‌یابد و صحت آن افزایـش می‌یابد. فرض کنید جمعیـت الگوریتم هریس هاکس ۱۰ و تعداد تکرارها ٤۰ و هر آزمایش ۳۰ بار تکرار شود. در این حالت میانگین خطای روش پیشنهادی حدود ۰/۰۹۸و دقت متوسط ۹۹/٦۸درصد است.

صحت روش پیشـنهادی از روش‌هـای دیگر، از جمله شـبکه‌های عصبی مصنوعی چندلایه، شـبکه‌های عصبی مصنوعی بازگشتی، ماشین‌های بردار پشتیبان و جنگل‌های تصادفی برای تشخیص فیشینگ فراتر می‌رود. این برتری هنگام مقایسـه روش پیشـنهادی با تحقیـق رائو [۲٤] در مورد انتخاب ویژگی و یادگیری ماشین مشـهود است. در این مقاله استفاده از روش‌های انتخاب ویژگی، مانند کاهش مولفه اصلی، در ترکیب با روش‌های یادگیری ماشین مانند جنگل تصادفی، درخت تصمیم‌گیری و شـبکه‌های عصبی مصنوعی چندلایه برای تشـخیص فیشینـگ مورد بحث قرار می‌گیرد. میانگین دقت جنگل تصادفی، درخت تصمیم و شبکه بیزی برای تشـخیص فیشینگ در این روش‌ها به ترتیب ۹۹/۳۱، ۹۸/۹۸ و ۹۸/۸۲ درصد است. در مقابل، دقت روش پیشـنهادی بالاتر اسـت، که اثربخشی آن را نشان می‌دهد و اعتماد به پتانسیل آن را القا می‌کند.

شکل ۹ نشان می‌دهد که روش پیشنهادی از روش‌های ترکیـبی انتخاب ویـژگی و یادگیری ماشیـن مانند جنگل تصادفی، درخت تصمیم و شبکه بیزی دقیق‌تر است. روش پیشنهادی حدود ۰/۳۷ درصد، ۰/۷ درصد و ۰/۸٦ درصد دقیق‌تر از جنگل تصادفی ترکیبی، درخت تصمیم و شـبکه بیزی است. از سوی دیگر، عملکرد روش پیشنهادی نیازی به زمان اجرای بالایی مانند کاهش مولفه ندارد، زیرا زمان اجـرای الگوریتم هریس هاکس کمتـر از روش‌هایی مانند PCA است.

۵- نتیجه گیری

حملات فیشینگ، حملات مخرب هستند. در این حملات، تله‌گذار، یـک وب‌گاه قانونی را جعل می‌کند و نشـانی آن را از طریق ایمیل یا رسـانه‌های اجتمـاعی برای کاربران ارسال می‌کند. حملات فیشینگ باعث خسارات قابل‌توجهی می‌شـود و سـالانه هزینه‌های زیادی برای شـرکت‌ها و مؤسسـات مالی به همـراه دارد. اسـتفاده از روش‌های یادگیری ماشینی در تشـخیص فیشینـگ نتایج خوبی به

همراه داشـــته است. با این حال، بحرانی‌ترین چالش در این روش‌ها تعداد زیاد ویژگی‌ها اســـت کـــه کاهش آنها نقش اســـاسی در دقت روش‌های تشـــخیص فیشینگ دارد. یکی از راه‌های موثر برای انتخاب بهترین ویژگی‌ها، استفاده از هستی‌شـــناسی صفحات وب است. این مقاله روشی مبتنی بر الگوریتم بهینه‌ســـازی هریس‌هاکس و هستی‌شـــناسی بـــرای یافتن ویژگی‌های بهینه صفحات وب و پیوندها ارائه می‌کند. روش پیشنهادی از ویژگی‌های بهینه برای آموزش شبکه‌عصبی مصنوعی اســـتفاده می‌کند. آزمایش‌ها نشان داده‌اند که روش پیشنهادی در تشخیص حملات دقیق‌تر از روش‌های یادگیری ماشینی مانند شبکه‌عصبی مصنوعی، درخت تصمیم و جنگل تصادفی اســـت. روش پیشـــنهادی دقیق‌تر از چندین روش انتخاب ویژگی، مانند PCA همراه با درخت تصمیم، جنگل تصادفی و شـــبکه بیزی است. روش پیشنهادی حدود ۰/۳۷ درصد، ۰/۷ درصد و ۰/۸۶ درصد دقیق‌تر از جنگل تصادفی ترکیبی، درخت تصمیم و شـــبکه بیزی اســـت. در تحقیقات آینده، مرحلـــه انتخاب ویژگی را با شـــبکه یادگیری عمیق CNN انجام خواهیم داد و مرحله طبقه‌بندی صفحه را با یادگیری عمیق LSTM اعمال خواهیم کرد.

تشکر و قدردانی

نویســـندگان تمایـــل دارنـــد از حمایت مالی دانشـــگاه بزرگمهر قائنات برای انجام این پژوهش با شماره قرارداد ۳۹۲۳۱ تقدیر و تشکر نمایند.

مراجع

1. Tanimu, J., Shiaeles, S. & Adda, M. 2024. “A Comparative Analysis of Feature Eliminator Methods to Improve Machine Learning Phishing Detection,” Journal of Data Science and Intelligent Systems, vol. 2, no. 2, pp. 87-99.
2. Elgharbi, S. E., Yahia, M. A. & Ouchani, S. 2024. “Online Phishing Detection: A Heuristic-Based Machine Learning Framework,” in 2024 13th Mediterranean Conference on Embedded Computing (MECO), IEEE, pp. 1-4.
3. Abdullah, A. A. & Hussein, S. A. 2024. “Detection and Mitigation Distribution Denial of Service Attack Based on Blockchain Concept,” technology, vol. 2, p. 4.
4. Guo, L.; Jing, S., Wei, L. & Zhao, C.2023. “Crossfire Attack Defense Method Based on Virtual Topology,” in 2023 19th International Conference on Mobility, Sensing and Networking (MSN), IEEE, pp. 341-350.
5. Smrithy Singh, V. & Rene Robin, C. 2023. “A Censorious Interpretation of Cyber Theft and its Footprints,” I-Manager’s Journal on Information Technology, vol. 12, no. 1.
6. Bensaoud, A., Kalita, J. & Bensaoud, M. 2024. “A survey of malware detection using deep learning,” Machine Learning With Applications, vol. 16, p. 100546.
7. Abroshan, H., Devos, J., Poels, G. & Laermans, E. 2021. “Phishing happens beyond technology: the effects of human behaviors and demographics on each step of a phishing process,” IEEE Access, vol. 9, pp. 44928-44949.
8. Odeh, A., Keshta, I. & Abdelfattah, E. 2021. “Machine learningtechniquesfor detection of website phishing: a review for promises and challenges,” in 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, pp. 0813-0818.
9. Sheng, S., Wardman, B., Warner, G., Cranor, L., Hong, J. & Zhang, C. 2009.”An empirical analysis of phishing blacklists,”.
10. Moody, G. D., Galletta, D. F. & Dunn, B. K. 2017. “Which phish get caught? An exploratory study of individuals′ susceptibility to phishing,” European Journal of Information Systems, vol. 26, no. 6, pp. 564-584.
11. Gandotra, E. & Gupta, D. 2021. “An efficient approach for phishing detection using machine learning,” in Multimedia Security: Springer, pp. 239-253.
12. Do, N. Q., Selamat, A., Krejcar, O., Herrera-Viedma, E. & Fujita, H. 2022. “Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions,” IEEE Access.
13. Wei, W., Ke, Q., Nowak, J., Korytkowski, M., Scherer, R. & Woźniak, M. 2020. “Accurate and fast URL phishing detector: a convolutional neural network approach,” Computer Networks, vol. 178, p. 107275.
14. Zhu, E., Chen, Z., Cui, J. & Zhong, H. 2022. “MOE/RF: A Novel Phishing Detection Model based on Revised Multi-Objective Evolution Optimization Algorithm and Random Forest,” IEEE Transactions on Network and Service Management.
15. Saxena, A., Sharma, N., Agarwal, P. & Barotia, R. 2021. “Phishing website prediction by using cuckoo search as a feature selection and random forest and BF-tree classifier as a classification method,” in Rising threats in expert applications and solutions: Springer, pp. 765-776.
16. Nagunwa, T., Naqvi, S., Fouad, S. & Shah, H. 2019. “A Framework of New Hybrid Features for Intelligent Detection of Zero Hour Phishing Websites,” in International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on EUropean Transnational Education (ICEUTE 2019),

Springer, pp. 36-46.
17. Hakim, Z. M. et al., 2021. “The Phishing Email Suspicion Test (PEST) a lab-based task for evaluating the cognitive mechanisms of phishing detection,” Behavior research methods, vol. 53, no. 3, pp. 1342-1352.
18. Saravanan, P. & Subramanian, S. 2020. “A framework for detecting phishing websites using GA based feature selection and ARTMAP based website classification,” Procedia Computer Science, vol. 171, pp. 1083-1092.
19. Wang, J. 2022. “An Improved Genetic Algorithm for Web Phishing Detection Feature Selection,” in 2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML), IEEE, pp. 130-134.
20. Zhu, E., Ju, Y., Chen, Z., Liu, F. & Fang, X. 2020. “DTOF-ANN: an artificial neural network phishing detection model based on decision tree and optimal features,” Applied Soft Computing, vol. 95, p. 106505.
21. Bozkir, A. S. & Aydos, M. 2020. “LogoSENSE: A companion HOG based logo detection scheme for phishing web page and E-mail brand recognition,” Computers & Security, vol. 95, p. 101855.
22. Aljofey A. et al., 2022. “An effective detection approach for phishing websites using URL and HTML features,” Scientific Reports, vol. 12, no. 1, pp. 1-19.
23. Minocha, S. & Singh, B. 2022. “A novel phishing detection system using binary modified equilibrium optimizer for feature selection,” Computers & Electrical Engineering, vol. 98, p. 107689.
24. Rao, R. S. & Pais, A. R. 2019. “Detection of phishing websites using an efficient feature-based machine learning framework,” Neural Computing and Applications, vol. 31, no. 8, pp. 3851-3873.
25. Yerima, S. Y. & Alzaylaee, M. K. 2020. “High accuracy phishing detection based on convolutional neural networks,” in 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS), IEEE, pp. 1-6.
26. Mashaleh, A. S., Ibrahim, N. F. B., Al-Betar, M. A., Mustafa, H. M. & Yaseen, Q. M. 2022. “Detecting Spam Email with Machine Learning Optimized with Harris Hawks optimizer (HHO) Algorithm,” Procedia Computer Science, vol. 201, pp. 659-664.
27. Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M. & Chen, H. 2019. “Harris hawks optimization: Algorithm and applications,” Future generation computer systems, vol. 97, pp. 849-872.
28. Rao, R. S. & Ali, S. T. 2015. “Phishshield: a desktop application to detect phishing webpages through heuristic approach,” Procedia Computer Science, vol. 54, pp. 147-156.
29. Volkamer, M., Renaud, K., Reinheimer, B. & Kunz, A. 2017. “User experiences of torpedo: Tooltip-powered phishing email detection,” Computers & Security, vol. 71, pp. 100-113.
30. Ollmann, G. 2004. “The phishing guide understanding & preventing phishing attacks,” NGS Software Insight Security Research.
31. Dua, D. & Graff, C. ???. UCI Machine Learning Repository [Online] Available: http://archive.ics.uci.edu/ml