پنجمین کنفرانس ملی مباحث نوین در کامپیوتر و فناوری اطلاعات
5th National Conference on Advanced Topics in Computer and Information Technology
۱۴ آذر ماه ۱۴۰۰

ارائه روشی جهت تشخیص و شناسایی ایمیل‌های اسپم با استفاده از الگوریتم گرگ خاکستری

رضا نیرومند[۱] – مرجان عبدیزدان[۲]

(۱) گروه کامپیوتر - دانشگاه آزاد اسلامی واحد ماهشهر
r.niromand@yahoo.com
(۲) گروه کامپیوتر – دانشگاه آزاد اسلامی واحد ماهشهر
abdeyazdan87@yahoo.com

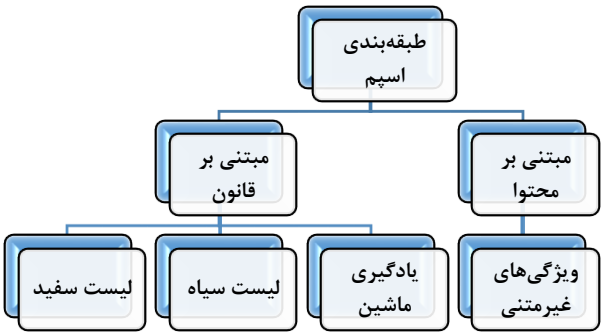
خلاصه: اسپم‌ها یا پست‌های الکترونیک ناخواسته که در سال‌های اخیر یکی از محبوب‌ترین روش‌های ارتباطی در بین مردم شناخته شده‌اند، پیام‌هایی با اهداف اقتصادی یا مخرب بوده که بدون رضایت گیرندگان پیام برای آنها ارسال می‌شوند. اسپم‌ها علاوه بر تاثیرات منفی که برای کاربران دارند تهدیدی امنیتی نیز به شمار می‌روند که باعث اختلال در کارایی سیستم‌های شبکه می‌شوند. برخی از آسیب‌هایی که توسط اسپم‌ها بوجود می‌آید شامل فیشینگ اطلاعات، کاهش پهنای باند، انتشار بدافزارها، افزایش محاسبات و هزینه بروزرسانی شبکه‌ها و تاثیر منفی بر سرویس‌های ارائه شده برای ایمیل‌های عادی می‌باشد. بنابراین باتوجه به مطالب بیان شده به روشی جهت تشخیص ایمیل‌های اسپم از عادی مورد نیاز است. یکی از روش‌های شناسایی و مقابله با ایمیل‌های اسپم استفاده از الگوریتم‌های یادگیری ماشین و هوش مصنوعی می‌باشد که در سال‌های اخیر مورد توجه بسیاری از محققان قرار گرفته است. از این‌رو با توجه به اهمیت تشخیص ایمیل‌های اسپم و توانایی الگوریتم‌های داده کاوی و هوش مصنوعی، در این تحقیق قصد داریم با استفاده از الگوریتم گرگ خاکستری روشی ارائه دهیم که دقت تشخیص و شناسایی ایمیل‌های اسپم از عادی را بهبود بدهد.

کلمات کلیدی: پست‌های الکترونیکی ناخواسته- اسپم-تشخیص-الگوریتم گرگ خاکستری.

۱ – مقدمه

امروزه در سراسر دنیا میلیون‌ها نفر هر روز برای ایجاد ارتباط با یکدیگر از ابزار ایمیل یا پست الکترونیک استفاده می‌کنند که یکی از قدیمی‌ترین و در عین حال متداول‌ترین سرویس‌های ارائه شده بر روی اینترنت می‌باشد[۱]. استفاده ارزان و راحت و زمان‌بر نبودن ارسال پیام بوسیله ایمیل یا پست الکترونیکی جهت اهداف مختلفی همچون علمی، تبلیغاتی، اجتماعی، فرهنگی و سیاسی طی دهه‌های اخیر سبب افزایش ارسال پیام در قالب پست الکترونیکی شده است[۲]. نامه‌های الکترونیکی ناخواسته‌ای که هر روز شاهد دریافت تعدادی از اینگونه نامه‌ها هستیم اسپم یا هرزنامه نامیده می‌شوند. با رشد فناوری اطلاعات و ارتباطات در فضای وب، وجود اسپم‌ها را می‌توان یکی از چالش‌های مهم فضای اینترنتی یا وب برشمرد[۳]. بنابراین در این راستا تکنیک‌های متعددی برای تشخیص اسپم‌ها مورد استفاده قرار می‌گیرد که یکی از رایج‌ترین این روش‌ها، روش مبتنی بر داده‌کاوی و هوش مصنوعی می‌باشد که طی سال‌های اخیر جهت تشخیص و مسدود اسپم‌ها بسیار مورد استفاده قرار گرفته می‌شوند [۵،۴].

جهت طبقه‌بندی پیام‌های وارد شده به صندوق پستی کاربران تکنیک‌های مختلفی وجود دارد که در شکل۱ یک نمونه از این طبقه‌بندی آمده است [۶] :



شکل ۱: تکنیک‌های طبقه‌بندی اسپم

با توجه به اینکه هرزنامه‌ها می‌توانند در انواع پیام‌رسان‌ها، گروه‌های خبری یوزنت، بخش نظرات وبلاگ‌ها نیز وارد شوند و بسته به اینکه چه منابعی را مورد حمله قرار خواهند داد، شامل دسته‌های مختلفی می‌باشند که عبارتند از [۸،۷]:

۱. نامه الکترونیکی
۲. گروه خبری
۳. شبکه اجتماعی
۴. تالار گفتمان اینترنتی
۵. وبلاگ
۶. پیام گوشی همراه
۷. صفحات اینترنتی

باتوجه به گستردگی مسئله اسپم در سیستم‌های نشانه‌گذاری اجتماعی، چالش‌های بسیار زیادی در این زمینه وجود دارد. در این سیستم‌ها اسپم نویسان در تولید اسپم می‌توانند اهداف گوناگونی داشته باشند. از جمله اهداف اسپم‌نویسان می‌توان به گمراهی کاربران، افزایش نرخ بازدید یا کاهش کیفیت موتور جستجو و کاهش عدم اعتماد کاربران به سیستم نشانه‌گذاری اشاره نمود. از این‌رو در سیستم‌های نشانه‌گذاری می‌بایست چالش‌های اسپم مورد بررسی قرار گرفته و راه‌های مقابله با آن اتخاذ گردد.

۲ – کارهای پیشین

در طول سال‌های گذشته، روش‌های متنوعی به منظور شناسایی اسپم انجام شده است که در ادامه به اختصار به چند مورد اشاره می‌شود:

- Asghar و همکاران در سال ۲۰۱۹ جهت تشخیص اسپم نظرات از طرح طبقه‌بندی ترکیبی استفاده نمودند. نتایج حاصل از این طرح نشان داد که استفاده از روش پیشنهادی آنها و انتخاب ویژگی‌های مربوط به اسپم سبب بهبود در شناسایی هرزنامه در سایت‌های مورد بررسی شده و با افزودن یک طرح وزندهی قادر خواهد بود ویژگی‌های اصلاح شده و بهینه‌تر را انتخاب نماید. بنابراین براساس روش پیشنهادی و نتایج بدست آمده می‌توان گفت دقت روش پیشنهادی از ۹۳ درصد به ۹۶ درصد افزایش یافته است[۹].
- Shuaib و همکاران در سال ۲۰۱۹ جهت تشخیص ایمیل‌های اسپم از الگوریتم نهنگ برای انتخاب بهترین ویژگی‌ها بر روی مجموعه داده SpamBase استفاده کردند. آنها پس از استفاده از الگوریتم نهنگ جهت انتخاب بهترین ویژگی‌ها و استفاده از الگوریتم جنگل تصادفی جهت طبقه‌بندی استفاده کردند. نتایج بدست آمده از روش پیشنهادی نشان داد که قادر است با دقت ۹۹,۸۶ درصد در حالت انتخاب ویژگی و دقت ۹۴,۲ درصد بدون انتخاب ویژگی به تشخیص ایمیل‌های اسپم بپردازد[۱۰].
- Kumaresan و همکاران در سال ۲۰۱۹ جهت تشخیص اسپم روشی با استفاده از تکنیک ماشین بردار پشتیبان و الگوریتم فاخته را ارائه دادند. در روش مذکور محققان از الگوریتم فاخته جهت بهبود کرنل ماشین بردار پشتیبان استفاده کردند که در نهایت توانستند با دقت ۹۴,۱۱ درصد به تشخیص ایمیل‌های اسپم بپردازند[۱۱].
- Azad و مورلا در سال ۲۰۱۸ به قابلیت فراهم کنندگان سرویس اینترنت در به اشتراک‌گذاری اطلاعات کاربران خود بین یکدیگر و تشخیص سریع ارسال کنندگان اسپم پرداختند. طبق رویکرد پیشنهادی بسته به میزان اطلاعاتی که بین فراهم کنندگان اینترنت به اشتراک گذاشته می‌شود ارسال کنندگان اسپم نیز با سرعت بالاتری مشخص و ردیابی خواهند شد. نتایج پیاده‌سازی نشان داده است رویکرد پیشنهادی این

مقاله کارایی بسیار بهتر در مقایسه با اطلاعات جزیره‌ای فراهم کنندگان دارد[۱۲].

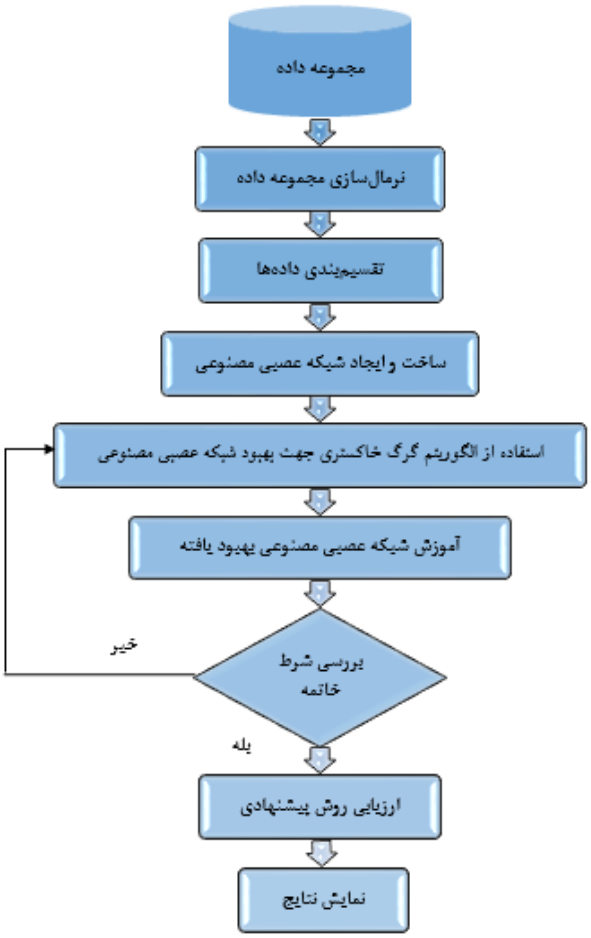
- Singh و همکاران در سال ۲۰۱۸ با استفاده از رویکرد آماری، احتمالی از ساختار اطلاعات سعی در تشخیص اسپم در اینترنت اشیا داشتند. آنها با مد نظر قرار دادن شبکه‌های اجتماعی فیسبوک، توئیتر و اینستاگرام، این شبکه‌ها را مورد تحلیل و بررسی قرار دادند. نتایج بدست آمده حاکی از عملکرد بالای روش پیشنهادی آنها است[۱۳].
- Vennila و همکاران در سال ۲۰۱۸ به بررسی اسپم‌های صوتی با استفاده از مدل مخفی مارکف پرداختند. این‌گونه از اسپم‌ها در اینترنت مزاحمت تلفنی محسوب شده و تفکیک آنها از تلفن‌های معتبر دارای اهمیت ویژه‌ای است. در واقع هدف اصلی تولیدکنندگان این‌گونه اسپم‌ها مصرف پهنای باند اینترنت و ایجاد مزاحمت محسوب می‌شود اگرچه گاهی پیام‌های تبلیغاتی نیز مدنظر هستند[۱۴].

۳ –فلوچارت روش پیشنهادی

همانطور که در فلوچارت روش پیشنهادی در شکل ۲ نشان داده شده است پس از دریافت مجموعه داده اسپم از سایت UCI، پیش‌پردازش بر روی داده‌های آن صورت گرفته و در ادامه داده‌ها تقسیم‌بندی می‌شوند. پس از تقسیم‌بندی داده‌ها شبکه عصبی مصنوعی ایجاد شده و تعداد لایه‌ها و نرون‌های میانی آن را توسط الگوریتم گرگ خاکستری تعیین خواهیم نمود. در نهایت پس از آموزش شبکه عصبی بهبود یافته بوسیله الگوریتم گرگ خاکستری و رسیدن به شرط پایان روش پیشنهادی را مورد بررسی و ارزیابی قرار خواهیم داد.

۴– پیاده‌سازی روش پیشنهادی

جهت پیاده‌سازی روش پیشنهادی، در این تحقیق از نرم‌افزار MATLABنسخه R2017a که یک زبان سطح بالا و با محیطی جذاب بوده و توسط شرکت MathWorks تولید شده است، استفاده گردید. این شرکت در سال ۱۹۸۴ در ایالت ماسوچوست امریکا تاسیس شد، در سال ۱۹۷۰ رئیس دانشگاه نیومکزیکو این نرم‌افزار را بر پایه زبان فرتون نوشته و پس از آن در سال ۱۹۸۳ این نرم‌فزار بر پایه زبان Cنوشته و گسترش پیدا کرد. واژه MATLAB که به معنای محیط محاسباتی رقمی و همچنین به معنای خود زبان برنامه‌نویسی مربوطه می‌باشد از ترکیب دو واژه ماتریس و آزمایشگاه ایجاد شده است. در واقع این نام حاکی از رویکرد ماتریس محور برنامه است که در آن حتی اعداد منفرد نیز به‌عنوان ماتریس در نظر گرفته می‌شوند. این نرم‌افزار با داشتن ویژگی‌های منحصر به فردی که دارد نسبت به سایر زبان‌های برنامه‌نویسی که از تمامی این ویژگی‌ها برخوردار نمی‌باشند کاربرد بیشتری در تمامی حوزه‌ها دارد.



شکل ۲: فلوچارت روش پیشنهادی

۵ –شرح روش پیشنهادی

در روش پیشنهادی پس از پیش‌پردازش بر روی مجموعه داده موردنظر جهت طبقه‌بندی داده‌ها ز الگوریتم‌های داده کاوی استفاده می‌شود. باتوجه به مجموعه داده موردنظر در این تحقیق پیش‌پردازش شامل نرمال‌سازی داده‌ها می‌باشد. برای این منظور در این مقاله برای نرمالیزه

کردن داده‌ها از روش نرمالیزاسیون آماری حداقل-حداکثر، در بازه [1,0] که از رابطه ۱ تبعیت می‌کند استفاده کردیم :

$$V_i' = \frac{v_i - min_A}{max_A - min_A}(new\ max_A - new\ min_A) + new\ min_A \quad (۱)$$

که در این رابطه $\min_A$ برابر با مینیمم مقدار ویژگی A، $\max_A$ برابر با ماکزیمم مقدار ویژگی A، new max_A برابر با ماکزیمم مقدار جدید ویژگی A، new min_A برابر با ماکزیمم مقدار جدید ویژگی A، v_i برابر با مقدار ویژگی A و V_i' برابر با مقدار ویژگی A بعد از نرمال‌سازی می‌باشد.

در گام بعدی جهت ارزیابی داده‌ها با استفاده از روش هولداوت داده‌ها به دو بخش آموزشی و ارزیابی تقسیم می‌شود(۷۰ درصد مجموعه داده برای آموزش و ۳۰ درصد آن برای آزمایش مورد استفاده قرار می‌گیرد).

پس از تقسیم‌بندی مجموعه داده، در گام بعدی شبکه عصبی مصنوعی از نوع پیشخور ایجاد می‌شود. پس از ایجاد شبکه عصبی پیشنهادی و تقسیم‌بندی داده‌ها، شبکه عصبی آموزش داده شد. در شبکه عصبی تعیین تعداد لایه‌های میانی و تعداد نرون‌های موجود در هر لایه میانی یکی از پارامترهای آزاد است که مقدار ثابت و مشخصی نداشته و برحسب سعی و خطا و تجربه کاربر تعیین می‌شود. لذا هدف از این تحقیق تعیین پارامترهای شبکه عصبی بوسیله الگوریتم گرگ خاکستری جهت افزایش دقت شبکه در تشخیص ایمیل‌های اسپم است.

۶- نگاشت مساله با استفاده از الگوریتم گرگ خاکستری

در روش پیشنهادی هر گرگ به صورت تصادفی در محیط قرار گرفته و برابر با ساختار یک شبکه عصبی است. نحوه چگونگی تعیین تعداد لایه‌ها و نرون‌های میانی بوسیله الگوریتم گرگ خاکستری بدین‌گونه است که هر گرگ به‌صورت برداری در جمعیت اولیه بوده که هر خانه از این بردار بیانگر تعداد لایه‌ها و نرون‌های میانی شبکه عصبی خواهد بود. لذا با توجه به جمعیت اولیه گرگ‌ها در این تحقیق، ماتریسی براساس تعداد جمعیت اولیه گرگ‌ها تشکیل می‌شود که تعداد ستون‌های آن برابر با تعداد لایه‌های میانی شبکه عصبی بوده و مقادیر موجود در هر خانه ماتریس نیز با تعداد نرون‌های لایه میانی برابر خواهد بود. پس از تعیین پارامترهای ورودی مسئله، روش پیشنهادی کار خود را با استفاده از داده‌های آموزشی و با یک لایه میانی آغاز می‌کند. بنابراین در ابتدا ماتریسی با یک ستون که نمایانگر یک لایه میانی می‌باشد و به تعداد ۳۰ سطر معادل جمعیت اولیه گرگ‌ها تشکیل می‌شود. در این حالت الگوریتم به تعداد ۱۰۰ مرتبه تکرار شده و در هر تکرار تمامی گرگ‌ها مقادیر متفاوتی نرون میانی خواهند داشت که به صورت تصادفی در محدوده بین ۱ تا ۵۰ می‌باشد. در هر تکرار میزان تابع هزینه هر گرگ نیز محاسبه می‌شود و در پایان تکرارها آن گرگی که نسبت به بقیه گرگ‌ها دارای تابع هزینه کمتری است به عنوان گرگ برتر انتخاب می‌گردد (گرگ آلفا).

در ادامه الگوریتم با دو لایه میانی اجرا می‌شود. بنابراین ماتریسی تشکیل می‌شود که دارای دو ستون و ۳۰ سطر بوده و به تعداد ۱۰۰ مرتبه تکرار می‌شود. در این حالت نیز الگوریتم به تعداد ۱۰۰ مرتبه تکرار شده و در هر تکرار تمامی گرگ‌ها به صورت تصادفی در محدوده بین ۱ تا ۵۰ نرون میانی انتخاب می‌کنند. در این حالت نیز میزان تابع هزینه برای همه گرگ‌ها محاسبه شده و بهترین گرگ انتخاب می‌شود.

در پایان الگوریتم کار خود را با سه لایه میانی آغاز می‌کند. یعنی ماتریسی تشکیل می‌شود که دارای سه ستون و ۳۰ سطر است و تمامی گرگ‌ها به صورت تصادفی در محدوده بین ۱ تا ۵۰ نرون میانی انتخاب می‌کنند. در این حالت نیز پس از ۱۰۰ مرتبه تکرار الگوریتم میزان تابع هزینه برای همه گرگ‌ها طبق رابطه ۲ محاسبه شده و بهترین گرگ که دارای کمترین تابع هزینه است انتخاب می‌شود.

$$\text{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \quad (۲)$$

که در این رابطه y_i برابر با خروجی واقعی، $\hat{y}_i$ برابر با خروجی پیش‌بینی شده توسط شبکه عصبی و n نیز برابر با تعداد کل فایل‌های آموزشی است.

پس از اینکه الگوریتم به تعداد لایه‌های میانی مورد نظر اجرا شد کار ارزیابی آن آغاز می‌گردد. در این مرحله از میان سه حالت اجرا شده الگوریتم مورد بررسی قرار می‌گیرد که در کدام حالت میزان تابع هزینه کمتر بوده است. بنابراین پس از مشخص شدن حالت موردنظر شبکه عصبی براساس تعداد لایه میانی و تعداد نرون میانی موجود در هر لایه یا لایه‌های میانی بوسیله مجموعه داده آزمایشی و با استفاده از معیارهای موردنظر مورد بررسی و ارزیابی قرار گرفته می‌شود.

۷- معیارهای ارزیابی روش پیشنهادی

در این تحقیق برای سنجش و ارزیابی عملکرد روش پیشنهادی براساس ماتریس درهم‌ریختگی از معیارهای دقت، صحت، فرخوانی و

F1 استفاده می‌شود.

$$Accuracy = (\frac{TP+TN}{TP+TN+FP+FN}) \times 100$$

$$Precision = (\frac{TP}{TP+FP}) \times 100$$

$$Recall = (\frac{TP}{TP+FN}) \times 100$$

$$F1 = \frac{2 \times Recall \times precision}{Recall + precision}$$

۸- نتایج پیاده‌سازی روش پیشنهادی

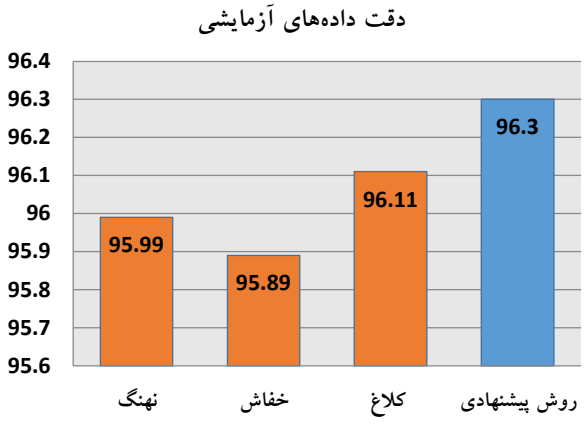
در این بخش به بررسی نتایج حاصل از پیاده‌سازی روش پیشنهادی باتوجه به معیارهای ارزیابی که در بخش قبل ذکر شد می‌پردازیم. برای این منظور از الگوریتم گرگ خاکستری جهت بهبود شبکه عصبی استفاده کردیم. نتایج بدست آمده برای داده‌های آموزشی و آزمایشی در جدول ۱ نشان داده شده است.

جدول ۱: نتایج بدست آمده برای داده‌های آموزشی و آزمایشی

	دقت	صحت	فراخوانی	F1
داده‌های آموزشی	۹۸,۷۵	۹۷,٤۷	۹۸,٦۲	۹۸,۰٤
داده‌های آزمایشی	۹٦,۳۰	۹٥,۸۲	۹۷,۸٦	۹٦,۸۲

نتایج بدست آمده از پیاده‌سازی روش پیشنهادی حاکی از آن است که روش پیشنهادی در این تحقیق از عملکرد و دقت بالایی در تشخیص ایمیل‌های اسپم برخوردار است.

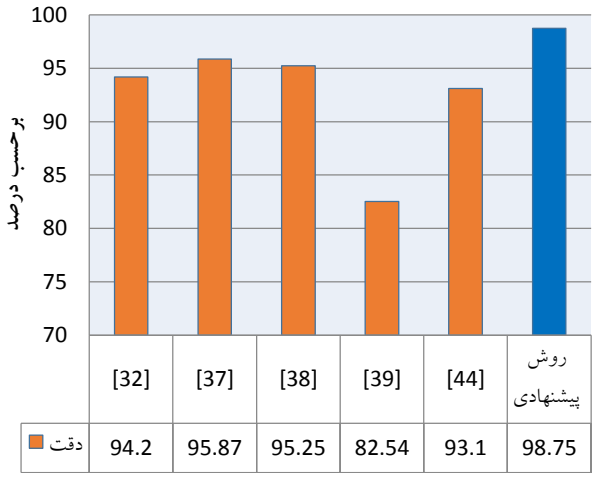
در ادامه به مقایسه دقت بدست آمده از روش پیشنهادی با برخی دیگر الگوریتم‌های فراابتکاری همچون الگوریتم‌های خفاش، نهنگ و کلاغ خواهیم پرداخت که در شکل۳ دقت روش پیشنهادی و الگوریتم‌های ذکر شده برای داده‌های آزمایشی نشان داده شده است.



شکل ۳: مقایسه دقت روش پیشنهادی با برخی از الگوریتم‌های فراابتکاری

مطابق با شکل ۳ می‌توان گفت دقت روش پیشنهادی با استفاده از الگوریتم گرگ‌خاکستری نسبت به الگوریتم کلاغ ۰,۱۹ درصد، نسبت به الگوریتم خفاش ۰,۴۱ درصد و نسبت به الگوریتم نهنگ ۰,۳۱ درصد بیشتر می‌باشد. بنابراین نتایج بدست آمده حاکی از آن است که روش پیشنهادی در این تحقیق از عملکرد خوب و بالایی جهت تشخیص ایمیل‌های اسپم برخوردار است.

در ادامه قصد داریم دقت بدست آمده از روش پیشنهادی در این تحقیق را با کارهای پیشینی که از مجموعه داده‌ای که ما در تحقیق خود از آن استفاده نمودیم، استفاده کرده‌اند مورد ارزیابی و مقایسه قرار دهیم. این مقایسه در شکل ۴ نشان داده شده است.



شکل ۴: مقایسه روش پیشنهادی با کارهای پیشین

۹- نتیجه گیری

نتایج بدست آمده از پیاده‌سازی حاکی از آن است که روش پیشنهادی با دقت ۹۸,۷۵ درصد برای داده‌های آموزشی و دقت ۹۷,۳۰ درصد برای داده‌های آزمایشی می‌تواند روش مناسبی جهت تشخیص ایمیل‌های اسپم باشد. همچنین باتوجه به نتایج بدست آمده و مقایسه آن با کارهای پیشین و همچنین الگوریتم‌های کلاغ، خفاش و نهنگ دریافتیم که روش پیشنهادی با عملکرد خوب و بالایی که دارد، می‌تواند به عنوان یک روش مناسب و کارامد برای تشخیص و دسته‌بندی ایمیل‌های اسپم از ایمیل‌های عادی مورد استفاده قرار بگیرد. در واقع می‌توان گفت علت عملکرد بهتر روش پیشنهادی تعیین تعداد مناسب لایه و نرون‌های میانی شبکه عصبی بوسیله الگوریتم گرگ خاکستری است که نقش مهمی در پردازش داده‌ها دارد؛ زیرا الگوریتم گرگ خاکستری با توانایی بالا در جستجوی نقاط بهینه و بهینه سراسری و با همگرایی بهتر و یادگیری قوی‌تر، موجب بهبود کارایی سیستم تشخیص ایمیل‌های اسپم از ایمیل‌های عادی شده است.

مراجع

[1] Ferrara, E. (2019). The history of digital spam. arXiv preprint arXiv:1908.06173.

[2] Ren, Y., & Ji, D. (2019). Learning to Detect Deceptive Opinion Spam: A Survey. IEEE Access, 7, 42934-42945.

[3] Broadhurst, R., & Trivedi, H. (2018). Malware in Spam Email: Trends in the 2016 Australian Spam Intelligence Data. Available at SSRN 3413442.

[4] Fang, Y., Zhang, C., Huang, C., Liu, L., & Yang, Y. (2019). Phishing Email Detection Using Improved RCNN Model With Multilevel Vectors and Attention Mechanism. IEEE Access, 7, 56329-56340.

[5] Salihovic, I., Serdarevic, H., & Kevric, J. (2018, June). The Role of Feature Selection in Machine Learning for Detection of Spam and Phishing Attacks. In International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies (pp. 476-483). Springer, Cham.

[6] Z. Liu, W. Lin, N. Li, and D. Lee, "Detecting and filtering instant messaging spam-a global and personalized approach," in *1st IEEE ICNP Workshop on Secure Network Protocols, 2015.(NPSec)*. 2015, pp. 19-24.

[7] J. Bi, J. Wu, and W. Zhang, "A trust and reputation based anti-spim method ",in *INFOCOM 2018. The 27th Conference on Computer Communications. IEEE*, 2018.

[8] Thakkar, A., & Lohiya, R. (2020). Role of swarm and evolutionary algorithms for intrusion detection system: A survey. Swarm and Evolutionary Computation, 53, 100631.

[9] Asghar, M. Z., Ullah, A., Ahmad, S., & Khan, A. (2019). Opinion spam detection framework using hybrid classification scheme. Soft Computing, 1-24.

[10] Shuaib, M., Adebayo, O. S., Osho, O., Idris, I., Alhassan, J. K., & Rana, N. (2019). Whale optimization algorithm-based email spam feature selection method using rotation forest algorithm for classification. SN Applied Sciences, 1(5), 390

[11] Kumaresan, T., Saravanakumar, S., & Balamurugan, R. (2019). Visual and textual features based email spam classification using S-Cuckoo search and hybrid kernel support vector machine. Cluster Computing, 22(1), 33-46.

[12] Azad, M. A., & Morla, R. (2019). Rapid detection of spammers through collaborative information sharing across multiple service providers. *Future Generation Computer Systems*, *95*, 841-854.

[13] Singh, A., & Batra, S. (2018). Ensemble based spam detection in social IoT using probabilistic data structures. *Future Generation Computer Systems*, *81*, 359-371.

[14] Vennila, G., Manikandan, M. S. K., & Suresh, M. N. (2018). Dynamic voice spammers detection using Hidden Markov Model for Voice over Internet Protocol network. *Computers & Security*, *73*, 1-16.