International Science and Technology Journal
المجلة الدولية للعلوم والتقنية
Volume 36 العدد
Part 1 المجلد
المجلة الدولية للعلوم والتقنية
International Science and Technology Journal
ISTJ

http://www.doi.org/10.62341/mksk1904

Received	2024/11/25	تم استلام الورقة العلمية في
Accepted	2024/12/22	تم قبول الورقة العلمية في
Published	2025/01/02	تم نشر الورقة العلمية في

دراسة أداء مُصنِّف مقترح لخوارزمية هجينة لإكتشاف التصيد الاحتيالي عبر البريد الإلكتروني

محمد خليفة صالح خليفة [1]

[1] قسم علوم الحاسب، كلية تقنية المعلومات، جامعة سبها – ليبيا

Email: moh.khalifa@sebhau.edu.ly

الملخص:

يهدف هذا البحث إلى دراسة فاعلية وأداء المُصنِّف المقترح لأكتشاف رسائل البريد الإلكتروني للتصيد الاحتيالي، وذلك لأن هناك حاجة مُلحة لتطوير أنظمة أمن معلومات قادرة بشكل دقيق واستباقي للتعرف على رسائل التصيد الاحتيالي بسبب عددها المتزايد وتنوع قدراتها الاحتيالية. نظرًا لأن هذا النوع من رسائل التصيد يتلاعب بالعواطف البشرية مما يؤدي إلى مخاوف ويخلق موقفًا وحالة عاجلة من خلال الادعاء بأن المستلم يجب أن يتخذ إجراءً فوريًا وسريعًا، مما قد يؤدي ويتسبب في خسائر مالية أوخسائر كبيرة في تسريب البيانات. من أجل التغلب على الضعف البشري فيما يتعلق باكتشاف رسائل البريد الإلكتروني المخادعة والتعرف عليها؛ تم إجراء هذه الدراسة، لأننا بحاجة إلى تعزيز وتحسين مستمر لدقة وفاعلية طرق وأساليب خوارزميات الكشف عن التصيد الاحتيالي بشكل آلي وتلقائيً. يستخدم نموذج التصنيف المقترح خوارزمية هجينة تجمع بين خوارزميات الشبكة العصبية متعددة الطبقات (MLP) للتعلم العميق (DL) وأساليب معالجة اللغة الطبيعية (NLP) عبر جسم النص الأساسي للبريد الإلكتروني المستلم. تسلط هذه الورقة الضوء على أهمية فحص الميزات النصية لجسم رسالة البريد لاكتشاف التصيد الاحتيالي، باستخدام الشبكات العصبية متعددة الطبقات (MLP) لتحسين دقة اكتشاف التصيد من خلال نص الرسالة، ولأن ميزات النص تمثل اتجاهًا جديدًا نسبيًا للدراسة في مجال اكتشاف التصيد عبر البريد الإلكتروني. تم اختبار النموذج المقترح باستخدام مجموعة بيانات متوازنة ومُصنَّفة تضم 8579 رسالة مختلفة، وأظهرت النتائج تحسنًا في دقة التصنيف والأداء مقارنة بأساليب التعلم العميق الأخرى. تم تقييم نموذج

http://www.doi.org/10.62341/mksk1904

المُصنِّف المقترح بأستخدام المقاييس التالية :(مقياس الاسترجاع "Recall"، معدل الدقة الشاملة "Accuracy"، معدل الانضباط للدقة "Precision"، ومقياس المتوسط التوافقي "f-measure")، وتم الحصول على النتائج _ بنسبة 98.3٪، 98.2٪، 98.5٪، و98.55٪ على التوالي. كما أظهر النموذج أيضًا أداءً جيدًا واستغرق وقتًا قصيرًا للكشف؛ لإنتاج معدل دقة شامل يزيد عن 98.1٪ ومعدل إيجابي زائف منخفض (FPR) قدره 0.015 .

الكلمات المفتاحية: كشف رسائل البريد الإلكتروني للتصيد الاحتيالي، الشبكة العصبية متعددة الطبقات "MLP"، التعلم العميق "DL"، معالجة اللغة الطبيعية "NLP"، الخوارزميات، أمن المعلومات.

Studying the Proposed Classifier Performance of a Hybrid Algorithm for E-mail Phishing Detection

Mohamed K. S. Khalifa [1]

[1] Computer Science Department, Faculty of Information Technology, Sebha University - Libya

Abstract :

This research aims to study the effectiveness and performance of the proposed classifier to detect phishing emails, because there is an urgent need to develop information security systems that are accurately and proactively able to recognize phishing messages due to their increasing number and diversity of fraudulent capabilities. Since this type of phishing message manipulates human emotions leading to fears and creates a situation of urgency by claiming that the recipient must take immediate and swift action, which may lead to financial losses or significant data leakage losses. In order to overcome the human weakness in detecting and recognizing phishing emails, this study was conducted because we need to continuously enhance and improve the accuracy and effectiveness of automated and automatic phishing detection algorithms and methods. The proposed classification model utilizes a hybrid algorithm that combines deep learning (DL) multi-layer perceptron (MLP) neural network algorithms and natural language processing (NLP) methods on the body of the received email. This paper highlights the importance of examining the textual features of the

http://www.doi.org/10.62341/mksk1904

body of a mail message for phishing detection, using multi-layer perceptron (MLP) neural networks to analyze the accuracy of detecting phishing through the message text, and because text features represent a relatively new direction of study in the field of email phishing detection. The proposed model was tested on a balanced and labeled dataset of 8579 different messages, and the results showed an improvement in classification accuracy and performance compared to other deep learning methods. The proposed classifier model was evaluated using the following metrics: (Recall, Accuracy, Precision, and F-measure), and the results were obtained - 98.3%, 98.2%, 98.5%, and 98.55%, respectively. The model also showed good performance and took a short time to detect; to produce an overall accuracy rate of over 98.1% and a low false positive rate (FPR) of 0.015.

Keywords: Phishing Email detection, MultiLayer Perceptron (MLP) Neural Network, Deep learning (DL), Natural language processing (NLP), Algorithms, Information security.

1. المقدمة :

خلال العقود الأخيرة، بلغ عدد الهجمات الإلكترونية الموجودة يوميًا، وفقًا لبعض الدراسات والإحصائيات إلى أكثر من 2200 هجوم إلكتروني يوميًا [1]؛ وهذا ما يعادل هجومًا إلكترونيًا واحدًا كل 39 ثانية. للأسف، ترافق التطور السريع والارتفاع المتزايد للخدمات من خلال الإنترنت مع تزايد عدد ومعدل الهجمات الإلكترونية، حيث يعد التصيد الاحتيالي عبر البريد الإلكتروني أحد أكثر الهجمات الإلكترونية شيوعًا وفعالية [2]. التصيد الاحتيالي هو نوع من الهجمات الإلكترونية التي تستخدم تقنيات الهندسة الاجتماعية والحيل التقنية للخداع والحصول على المعلومات والبيانات الحساسة من المستخدمين وسرقتها [3]، [4]، [5]. يتم ذلك عن طريق استغلال نقاط الضعف والثغرات الموجودة في عمليات النظام من جانب المستخدم النهائي؛ تعتمد مكافحتها على المعرفة والوعي البشري، والتي يمكن أن تختلف وتتفاوت بين المستخدمين. حتى إذا كان النظام آمنًا بشكل كافٍ ضد الجرائم الإلكترونية، يمكن للمهاجم أن يخدع مستخدمًا غافلًا للكشف عن بيانات الهوية الشخصية أو بيانات اعتماد حسابه المالي المصرفي. والأسوأ من ذلك، يمكن للمهاجم أن يعرض المؤسسات للخطر عن طريق تضليل أو إستغفال أحد الموظفين المسؤولين لتسهيل الوصول إلى بيانات المنظمة أو قواعد البيانات أو غيرها من

http://www.doi.org/10.62341/mksk1904

المعلومات الحساسة بالمؤسسة. وفقًا لـ شركة أمن الكمبيوتر والبرمجيات الأمريكية مكافي "McAfee"، تقدر الخسائر العالمية السنوية للجرائم الإلكترونية التي حسبتها شركة مكافي "McAfee" بحوالي 1 تريليون دولار أمريكي [6]. علاوة على ذلك، يعتبر التصيد الاحتيالي جريمة إلكترونية خطيرة وواسعة الانتشار تؤثر على كل من الشركات والمؤسسات والأفراد [7].

وفقًا لتقرير نُشر عن التصيد الاحتيالي، تعرضت 75٪ من المؤسسات والشركات لهجوم تصيد في عام 2021، ووصلت 96٪ من هجمات التصيد عبر البريد الإلكتروني [8]. من بين جميع الخروقات الأمنية، كان 22٪ متعلقًا بالتصيد الاحتيالي، وفقًا لتقرير التحقيقات في خرق البيانات لعام 2020 الصادر عن شركة الاتصالات اللاسلكية الأمريكية فيريزون "Verizon" [9]. أصبحت المنظمات والشركات بشكل متزايد أهدافًا للمهاجمين الذين يحاولون سرقة الأموال، أو الأسوأ من ذلك، بياناتهم القيمة. إن عواقب سرقة البيانات وخيمة وخطيرة: فهي تلحق الضرر بسمعة وقيمة المؤسسات والشركات. علاوة على ذلك، فإنه يتسبب في خسارة العملاء للقطاعات التي لديها بيانات حساسة، مثل قطاعي البنوك والاتصالات. لذلك، يجب على الشركات والمؤسسات اعتماد وتبنى آلية قوية للكشف عن التصيد الاحتيالي.

يعتبر أسلوب التصيد عبر رسائل البريد الإلكتروني المخادعة من أكثر تقنيات التصيد الاحتيالي شيوعًا [8]، [10]؛ حيث يستخدم المهاجم بريدًا إلكترونيًا مزيفًا لخداع المستلم لإرسال معلوماته الحساسة. هناك العديد من الطرق لإقناع المستلم بالكشف عن معلوماته الحساسة، مثل (النقر فوق رابط وزيارة موقع ويب مزيف لإدخال اسم المستخدم وكلمة المرور الخاصة به، أو إرسال المعلومات مباشرةً، أو تنزيل مرفق بريد إلكتروني يقوم بتنفيذ برامج تجسس ضارة). إن العدد المتزايد من رسائل البريد الإلكتروني المخادعة ينذر بالخطر، وفقًا لتقرير مجموعة عمل مكافحة التصيد الاحتيالي "APWG"؛ نما وارتفع متوسط تكلفة الاعتداءات وخسائر هجمات التصيد عبر رسائل البريد الإلكتروني في الربع الثالث من عام 2020 للشركات والمؤسسات من 48,000 $ ألف دولار إلى 75,000 $ ألف دولار [11]. علاوة على ذلك، ارتفع عدد رسائل التصيد الإلكتروني المسجلة على مستوى العالم في الربع الأخير من عام 2020 بنسبة 18٪ مقارنة بالربع الثاني من عام 2019 [11]. بالإضافة إلى ذلك، في عام 2020 وفقًا لتقرير الدفاع

http://www.doi.org/10.62341/mksk1904

الرقمي لشركة مايكروسوفت "Microsoft"، بلغ عدد رسائل البريد الإلكتروني الضارة في برنامج "Microsoft 365" إلى 13 مليار رسالة [12].

وفقًا لـ بينسون وآخرون "Benenson et al" [13]، كشفت إحدى تجارب التصيد الاحتيالي أن 20٪ من الجمهور استجابوا وقاموا بالنقر فوق الرابط المزيف المضمّن في رسائل البريد الإلكتروني؛ عندما سُئلوا عن سبب قيامهم بالنقر على الرابط، ذكر 34٪ ممن شملهم الاستطلاع أنه كان بدافع الفضول. وأوصى هؤلاء الباحثون بأن تقوم الشركات والمؤسسات بحماية الموظفين، قدر الإمكان، من عملية اتخاذ القرار بفتح رسائل البريد الإلكتروني المخادعة والرد عليها. بناء على ذلك، فإن وجود تقنيات وخوارزميات آلية للكشف تلقائيًا عن رسائل البريد الإلكتروني المخادعة باستخدام فحص النص الأساسي للبريد الإلكتروني المستلم يعد أمرًا ضروريًا لمكافحة هذه المخاطر والتهديدات المتزايدة.

على الرغم من حقيقة أن لدينا العديد من الأجهزة والأدوات للكشف عن رسائل البريد الإلكتروني المخادعة، إلا أن الشركات والمؤسسات والأفراد لا يزالون يعانون من عمليات التصيد الاحتيالي عبر البريد الإلكتروني. تتسم طبيعة رسائل البريد الإلكتروني التصيدية بأنها متغيرة وديناميكية ومرنة لأنها تعتمد على الذكاء البشري للمهاجم. بمرور الوقت، يقوم المهاجمون بتطوير وتغيير أساليب التصيد الخاصة بهم؛ وبالتالي، يجب مواجهة هذه المشكلة ومعالجتها من خلال إيجاد وابتكار حل يتناسب ويتوافق مع الذكاء البشري.

للتعامل مع هذه المشكلة ومعالجتها، تم إنشاء طرق وأساليب جديدة مثل التعلم العميق (DL)، والتي يمكن أن تنافس الذكاء البشري، ولكن هذه الأساليب لا تزال في مهدها وتحتاج إلى تحسين [1]. علاوة على ذلك، تعتمد معظم تقنيات التصيد الآلي الحالية للكشف عن رسائل البريد الإلكتروني الاحتيالية لتقليل الأخطاء البشرية باستخدام البيانات الوصفية للبريد الإلكتروني أو عناوين "URLs" المرفقة.

أظهرت الأبحاث السابقة لأكتشاف رسائل البريد الإلكتروني المخادعة نتائج واعدة، ولكن هناك حاجة إلى مزيد من العمل على نماذج أكتشاف رسائل التصيد الاحتيالي؛ نظرًا للطبيعة الديناميكية والمتغيرة لرسائل البريد الإلكتروني المخادعة، الأمر الذي يتطلب أساليب متقدمة. في الواقع، تم اقتراح عدد قليل جداً من نماذج اكتشاف التصيد المبنية

http://www.doi.org/10.62341/mksk1904

على تحليل البريد الإلكتروني [10]، وهناك حاجة ماسة إلى مزيد من العمل والجهد لتحسين دقة أكتشاف التصيد. بالإضافة إلى ذلك، أثبتت تقنية الشبكة العصبية متعددة الطبقات (MLP) نجاحها في تصنيف المستندات [14]، ولكن لم يتم حتى الآن تجربة أو اقتراح أي مُصنِّف بريد إلكتروني للتصيد الاحتيالي يعتمد على تحليل النص الأساسي باستخدام أساليب خوارزمية الشبكة العصبية متعددة الطبقات (MLP). هذه الدراسة هي من بين أولى الدراسات، على حد علمنا، التي تبحث وتتحقق من فعالية وكفاءة استخدام أساليب خوارزمية الشبكة العصبية متعددة الطبقات (MLP) لأكتشاف التصيد الاحتيالي في النص الأساسي لرسائل البريد الإلكتروني من أجل زيادة تحسين دقة اكتشاف التصيد. علاوة على ذلك، يمكن للمصنف أو الخوارزمية الهجينة المقترحة لهذه الورقة البحثية اكتشاف التصيد الاحتيالي من رسائل البريد الإلكتروني المستلمة حتى إذا كان البريد الإلكتروني لا يحتوي على عنوان "URL" أو مرفقات.

في البداية، تبدأ الورقة بمقدمة بينما يتم تنظيم بقية هذه الورقة على النحو التالي: يقدم القسم 2 نظرة عامة على الخلفية البحثية حول بعض التقنيات والأساليب المتعلقة باكتشاف البريد الإلكتروني المخادع، يليها القسم 3 الذي يصف المنهجية المستخدمة في هذه الدراسة. يغطي القسم الرابع ويصف بيئة نموذج المُصنِّف المقترح. بعد ذلك في القسم 5، يتم عرض الأدوات والنتائج والمناقشات، بالإضافة إلى اختبار وتقييم المصنف. أخيرًا، يتم تقديم الاستنتاجات مع لمحة عامة عن العمل المستقبلي للمشروع، ويليه الخلاصة والتوصيات.

2. نظرة عامة على خلفية البحث :

يعد البريد الإلكتروني للتصيد الاحتيالي أحد أسرع الجرائم الإلكترونية نموًا وانتشارًا على الإنترنت. الهدف من البريد الإلكتروني للتصيد الاحتيالي هو سرقة المعلومات السرية والخاصة للمستخدم من خلال التظاهر بأنه مصدر شرعي وموثوق. بالنسبة للمجرمين والمحتالين عبر الإنترنت، تطور التصيد الاحتيالي إلى عمل ونشاط تجاري مربح ومجدي. يمكن أن تتسبب هجمات التصيد الاحتيالي الناجحة في خسارة مالية للضحايا وتعرض أمن معلوماتهم وبياناتهم الشخصية للخطر. حتى أنه يؤثر على التقنيات الناشئة، بما في ذلك إنترنت الأشياء "Internet-of-Things" والخدمات السحابية "Cloud Services" [15]، [16].

http://www.doi.org/10.62341/mksk1904

هناك نوعان رئيسيان من الأساليب التقنية لاكتشاف رسائل البريد الإلكتروني المخادعة: القائمة السوداء "Blacklisting" والتعلم الآلي (ML). تقارن طريقة القائمة السوداء بين عنوان البريد الإلكتروني للمرسل أو عنوان بروتوكول الإنترنت "IP" أو عنوان نظام اسم المجال "DNS" مع قائمة محددة مسبقًا من عناوين التصيد الاحتيالي، وإذا كانت البيانات متطابقة، فسيتم رفض البريد الإلكتروني قبل أن يصل إلى خادم بريد "SMTP" [3]، [10]. في حين أن طريقة القائمة السوداء لها معدل إيجابي زائف منخفض، إلا أنها تعتمد بشكل أساسي على الإبلاغ عن رسائل البريد الإلكتروني المخادعة من جانب المستلم [5]. وعلى نفس المنوال، فإن القائمة البيضاء التلقائية تراقب رسائل البريد الإلكتروني الواردة والصادرة وتنتج مجموعة من العناوين الشرعية والموثوقة. كما تقوم أيضًا بمراجعة سجلات اتصالات البريد الإلكتروني السابقة المخزنة داخل خادم بريد إلكتروني للعناوين المراد إضافتها إلى القائمة. على الرغم من حقيقة أنه يمكن استخدام القائمة البيضاء لمنع أو إيقاف رسائل البريد الإلكتروني المخادعة، إلا أنها ليست فعالة بما يكفي لاكتشاف جميع هجمات التصيد الاحتيالي [17]. نظرًا لأن العناوين أو المواقع الجديدة لا يمكن تحديدها أو اكتشافها، فإنها لا توفر الحماية أو الأمان ضد الهجمات بشكل نهائي.

بدلاً من ذلك، يمكن للتعلم الآلي (ML) أتمتة (جعلها تعمل بشكل آلي) للكشف عن رسائل البريد الإلكتروني المخادعة من خلال طرق مختلفة، مثل اساليب الكشف للتعلم العميق (DL) التي تعمل تلقائيًا على اكتشاف البريد العشوائي (spam). لتحسين اكتشاف التصيد الاحتيالي، اقترح الباحثون طريقتين لتحسين المصنف:

1) أختبار/ تقييم عدة خوارزميات مختلفة .
2) التركيز على اختيار وتحديد الميزة بعناية لتحسين المصنف.

يُعرَّف اختيار الميزة على أنه طريقة للحصول على مجموعة فرعية من مجموعة ميزات أصلية تتطابق مع مقاييس اختيار ميزة معينة. من خلال هذه العملية، يتم تحديد الميزات المفيدة لمجموعة البيانات؛ ويساعد في تقليل حجم معالجة البيانات عن طريق إزالة الميزات غير الضرورية وغير ذات الصلة. يمكن أن يؤدي الاختيار الجيد للميزات إلى زيادة دقة التعلم الآلي (ML) [17] ، وتبسيط نتائج التعلم، وتقليل وقت التعلم [18]. يمكن استخدام ثلاث فئات من الميزات الرئيسية في اكتشاف رسائل البريد الإلكتروني المخادعة، كما هو موضح:

http://www.doi.org/10.62341/mksk1904

- فئة ميزة الرأس– Header Feature Class برسالة البريد الإلكتروني والتي تتضمن: (الطابع الزمني، خادم المستلم ، تنسيق المحتوى، ومعلومات الرأس الأخرى)؛
- فئة ميزة الجسم– Body Feature Class بالرسالة والتي تحتوي على فئتين فرعيتين: (ميزات النص ، وفئات عناوين URL)؛
- فئة المرفقات – Attachments Class.

التعلم العميق (DL) هو أحد فروع التعلم الآلي (ML) الذي يستخدم الخوارزميات لاستكشاف البيانات ذات الصلة أو المرتبطة ووضع نموذج لها، حيث يُمكّن التعلم العميق نموذج الكمبيوتر من تعلم مهام التصنيف وتنفيذها مباشرةً باستخدام الشبكات العصبية "NNs" ومجموعات البيانات "Datasets" التي تتضمن النصوص أو الأصوات أو الصور. في هذه الدراسة، السبب في اختيار التعلم العميق (DL) على خوارزميات التعلم الآلي (ML) الأخرى هو العمل المكثف لهندسة الميزات المطلوبة والمستخدمة لتنفيذ خوارزميات التعلم الآلي التقليدية وقابليتها للتطوير [5]، [19]. علاوة على ذلك، نظرًا لأن البيانات الموجودة في رسائل البريد الإلكتروني غير منظمة، فإن خوارزميات التعلم العميق (DL) هي الخيار الأفضل لهذا العمل [2]. بالإضافة إلى ذلك، أظهرت النتائج التي تم عرضها في الدراسات [2]، [5]، [19]، [20] فعالية تطبيق خوارزميات التعلم العميق (DL) في اكتشاف رسائل البريد العشوائي.

الشبكة العصبية متعددة الطبقات (MLP) هي خوارزمية التعلم العميق (DL) الموصوفة كنوع من الشبكات العصبية التلافيفية (CNN) التي تعمل مباشرة على الرسوم البيانية باستخدام البيانات المُنظّمة من خلال التعلم شبه الخاضع للإشراف لتصنيف العُقد "nodes". تُستخدم خوارزمية الشبكة العصبية متعددة الطبقات (MLP) بشكل شائع في مشاكل التصنيف مثل تصنيف المستندات / الوثائق؛ إنها ببساطة تحول مشكلة تصنيف الوثيقة / المستند إلى مشكلة تصنيف العُقدة "node". الفكرة وراء خوارزمية شبكة (MLP): أولاً، هي إنشاء الرسم البياني للمجموعة ببناء رسم بياني كبير واحد من مجموعة بريد إلكتروني كاملة، بحيث يتم تمثيل كلمات البريد الإلكتروني "email words" وجسم رسائل البريد الإلكتروني كعُقد "nodes"؛ ثم يتم بعد ذلك، إدخال وتغذية الرسم البياني في شبكة متعددة الطبقات.

في هذا المشروع البحثي، قمت بتصميم خوارزمية هجينة تجمع بين أساليب معالجة اللغة الطبيعية (NLP) لميزات نص البريد الإلكتروني وخوارزميات التعلم العميق باستخدام

http://www.doi.org/10.62341/mksk1904

الشبكة العصبية متعددة الطبقات (MLP) لبناء مصنف فعال لزيادة تحسين دقة الاكتشاف التلقائي للرسائل المخادعة للتصيد الاحتيالي عبر البريد الإلكتروني.

3. منهجية الدراسة:

نظراً لتزايد عدد وأساليب الرسائل الاحتيالية، ظهرت الحاجة إلى التطوير المستمر للأنظمة القادرة على الكشف عنها بشكل تلقائي ودقيق. في هذه الدراسة، أستخدمنا نص الرسالة للبريد الإلكتروني لتحسين دقة اكتشاف التصيد، حيث نعتقد أن هناك قدرًا كبيرًا من المعلومات القيمة مخبأة أو مخفية ضمن جسم نص رسالة البريد الإلكتروني تستحق الجهد المبذول للتدقيق. بناءً على ذلك، نقترح مُصنِّفًا جديدًا (خوارزمية هجينة) لاكتشاف البريد الإلكتروني للتصيد الاحتيالي استنادًا إلى خوارزميات التعلم العميق (DL) باستخدام الشبكة العصبية متعددة الطبقات (MLP) وأساليب معالجة اللغة الطبيعية (NLP). يحدد المصنف ما إذا كان البريد الإلكتروني شرعيًا أم أنه محاولة تصيد. للقيام بذلك، يتم تحليل جسم رسائل البريد الإلكتروني باستخدام خوارزميات التعلم العميق (DL) وتقنيات معالجة اللغة الطبيعية (NLP). ثم يتم إجراء تقييم لدقة المصنف بأستخدام المقاييس التالية: (الانضباط "Precision"، الدقة "Accuracy"، والاسترجاع "Recall") ومقارنتها بأحدث النماذج المنشورة. يعمل المصنف المقترح في هذا البحث بفاعلية وبشكل جيد على مجموعة بيانات متوازنة ومُصنَّفة؛ تتجاوز الدقة 98٪، ويتنافس ويتفوق على معظم مقاييس وإجراءات التقييم لنماذج وخوارزميات الكشف الحالية.

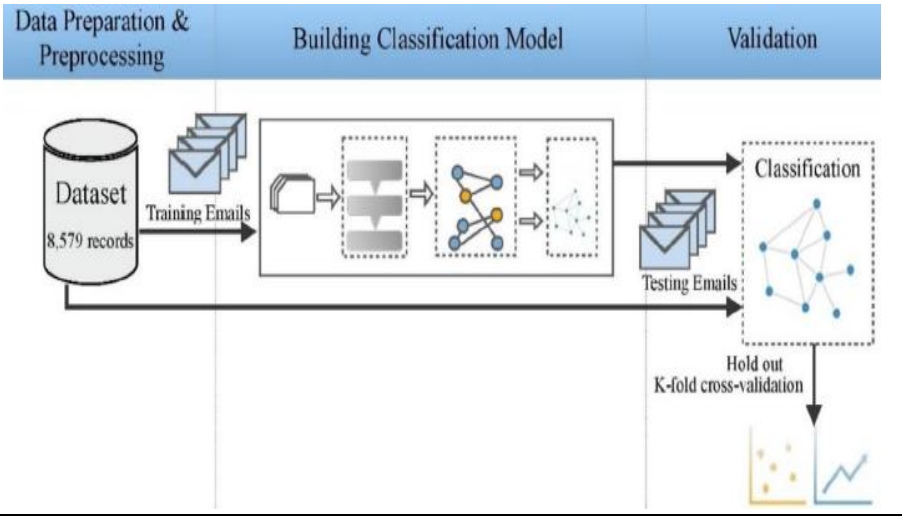
تم إجراء التجارب / الاختبارات المقترحة وتنفيذها على خادم البريد لاكتشاف رسائل البريد الإلكتروني المخادعة. أضف على ذلك، يمكن أن يكون لها تطبيقات أخرى في اكتشاف التصيد الاحتيالي من خلال تحليل النص في العديد من وسائل التصيد الشائعة مثل: رسائل الوسائط الاجتماعية أو المنشورات أو التغريدات.

4. نموذج المُصنِّف المقترح لاكتشاف التصيد الاحتيالي:

لبناء المصنف المقترح، فإننا نمر بثلاث مراحل رئيسية. أولاً، يتم جمع البيانات من رسائل التصيد الاحتيالي ورسائل المشروعة وإعدادها للتدريب والاختبار، كما هو مذكور وموضح في القسم 4.1 . تم إنشاء نموذج الاكتشاف المقترح من خلال خوارزميات الشبكة

http://www.doi.org/10.62341/mksk1904

العصبية متعددة الطبقات (MLP) للتعلم العميق، على النحو المبين في القسم 4.3 . يتم إدخال وتغذية بيانات التدريب إلى المصنف لبناء النموذج وإعداده للتقييم. بعد ذلك، تتمثل المرحلة الأخيرة في اختبار المصنف المقترح بطريقة خاضعة للإشراف باستخدام بيانات الاختبار للتحقق من صحة النموذج. يظهر "الشكل –1"، رسمًا بيانيًا وتخطيطيًا يصف سير العمل في إجراء عملية البحث.



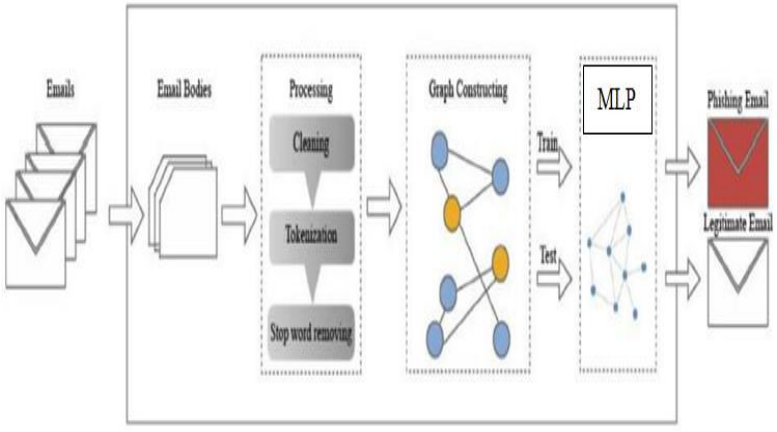
الشكل 1: المراحل الرئيسية لسير عمل البحث.

يعد اكتشاف رسائل البريد الإلكتروني المخادعة مشكلة ثنائية التصنيف، حيث يجب على المصنف التمييز بين نوعين من رسائل البريد الإلكتروني: (رسائل التصيد الاحتيالي، ورسائل البريد الإلكتروني الشرعية). بالنسبة للمدخلات الخاصة بمهمة التصيد للتصنيف، تكون التغذية / الإدخال عبارة عن مجموعة من الحالات (رسائل البريد الإلكتروني). كل حالة، المعروف أيضًا باسم السجل "record"، يتم تمثيله بواسطة مجموعة (x، y)، حيث x هي مجموعة السمات و Y هي تسمية الفئة أو سمة الهدف. إنها مهمة تعلم خاضعة للإشراف لأن أهداف المخرجات قد تم تحديدها مسبقًا. الهدف من مهمة التصنيف هو التنبؤ / التوقع بفئة عينة غير موسومة أو غير مصنفة. إذا كان لدينا مجموعة من الفئة المحدودة [C = {c1, c2,c3,.....}] ، يتم تعيين أو تخصيص العينة غير الموسومة x إلى فئة واحدة في C. في حالة اكتشاف رسائل البريد الإلكتروني المخادعة، يكون

http://www.doi.org/10.62341/mksk1904

التصنيف بمثابة تمثيل للبريد الإلكتروني مع تسمية أو تصنيف المجموعة المضبوطة على { احتيالي "phishing" ،أو شرعي "legitimate" }.

بمزيد من التفصيل، تبدأ عملية التصنيف بتغذية المصنف / النموذج المقترح برسائل البريد الإلكتروني التي تتم معالجتها لاستخراج النص الأساسي للرسالة. يتم تفكيكها وتنظيفها من الضوضاء والمعلومات غير ذات الصلة، وثم ترميزها، وإعدادها لإنشاء الرسم البياني. تم تدريب خوارزمية الشبكة العصبية متعددة الطبقات (MLP) على تصنيف وتمييز رسائل البريد الإلكتروني المخادعة / الاحتيالية، وغير المرغوب فيها. يوضح "الشكل –2"، نموذج مصنف الكشف المقترح المستخدم في هذا البحث، ويتم توفير وصف لكل خطوة موصوفة في القسم الفرعي التالي.



الشكل 2: نموذج تصنيف اكتشاف رسائل البريد الإلكتروني الاحتيالية.

4.1. المعالجة المسبقة للبيانات:

تحتاج البيانات إلى التنظيف لاستبعاد وإزالة الأحرف والكلمات غير الضرورية والتي لا داعي لها. بعد أن يتم تنظيف البيانات، يمكننا استخدام تقنيات استخراج الميزات / السمات الرسمية والأساسية. لذلك، يتضمن إعداد بيانات البريد الإلكتروني ما يلي:

1. استخراج أو استخلاص النص الأساسي من الجسم الرسالة.
2. التخلص وإزالة المسافات البيضاء الناتجة عن تحليل وتفسير النص.

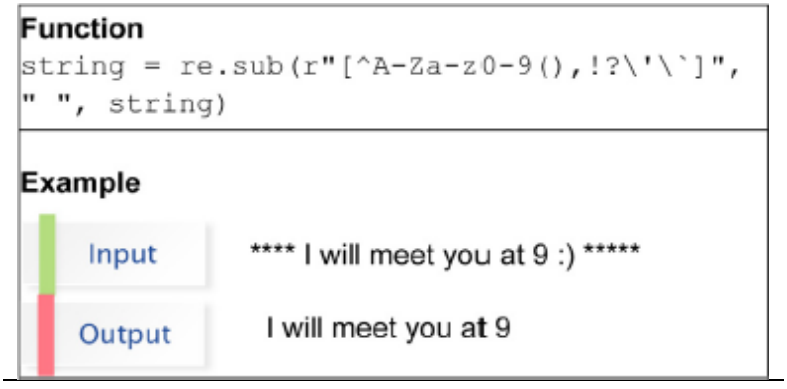
http://www.doi.org/10.62341/mksk1904

3. كتابة وتغيير جميع الأحرف إلى أحرف صغيرة وإزالة الأحرف غير الأبجدية الرقمية.

طبقت هذه الدراسة مختلف تقنيات المعالجة المسبقة والمتنوعة على رسائل البريد الإلكتروني، بما في ذلك (تنظيف البريد الإلكتروني، الترميز "tokenization"، وإزالة / تصفية كلمات التوقف والكلمات النادرة). يتم وصف كل منها على النحو التالي:

4.1.1. تنظيف البريد الإلكتروني:

تعمل هذه العملية على تنظيف رسائل البريد الإلكتروني من المعلومات غير ذات الصلة والأحرف غير الضرورية. تتم إزالة الأحرف غير الأبجدية الرقمية والأحرف غير الخاصة مثل ("؟"، "!"، و" ’ ")، باستخدام التعبير لـ "Python RegEx": وهو عبارة عن سلسلة من الأحرف التي تشكل نمط بحث؛ يمكن استخدام "RegEx (Regular Expression)" للتحقق مما إذا كانت السلسلة تحتوي على نمط البحث المحدد أم لا. كما تم أيضًا إزالة واستبعاد المساحات البيضاء. يعرض "الشكل –3" مثالاً على استخدام الرموز الزائفة وكيفية التخلص من البيانات غير ذات الصلة.

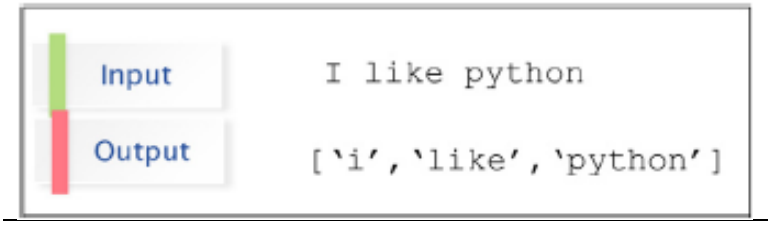


الشكل 3: تنظيف البريد الإلكتروني.

4.1.2. الترميز "Tokenization" :

تقسم عملية الترميز (tokenization) كل بريد إلكتروني إلى كلمات بناءً على المسافات البيضاء. تتم الإشارة إلى هذه الكلمات باسم الرموز المميزة "tokens". في هذه الدراسة، نستخدم وظيفة الدالة "a split() function" لتقسيم الجمل وترميزها. يُظهر "الشكل –4" مثالاً على عملية الترميز.

http://www.doi.org/10.62341/mksk1904



الشكل 4: ترميز البريد الإلكتروني.

4.1.3. إزالة كلمات التوقف والكلمات النادرة :

كلمات التوقف هي كلمات شائعة جدًا؛ عادة ما تكون كلمات تساعد في بناء الأفكار ولكنها لا تحمل في حد ذاتها أي دلالة أو أي أهمية، مثل: (حروف العطف، أدوات التعريف، حروف الجر، ... وما إلى ذلك). لقد استخدمنا قائمة كلمات التوقف من مجموعة أدوات اللغة الطبيعية (NLTK). تحتوي قائمة كلمات الإيقاف على عبارات مثل: ("off"، "no"، "aren't"، "too"، "an"، "being"، "only"، "ll"، "o"، "its"، "them"، "mightn't"، وغيرها ... إلخ). يعرض "الشكل –5" مثالاً على البنية البرمجية لعملية إزالة كلمات التوقف والكلمات النادرة .

```
For each word in words:

  IF word not in stop_words and
word_freq[word] >= 7:

Email_words.append(word)
```

الشكل 5: إزالة كلمات التوقف والكلمات النادرة.

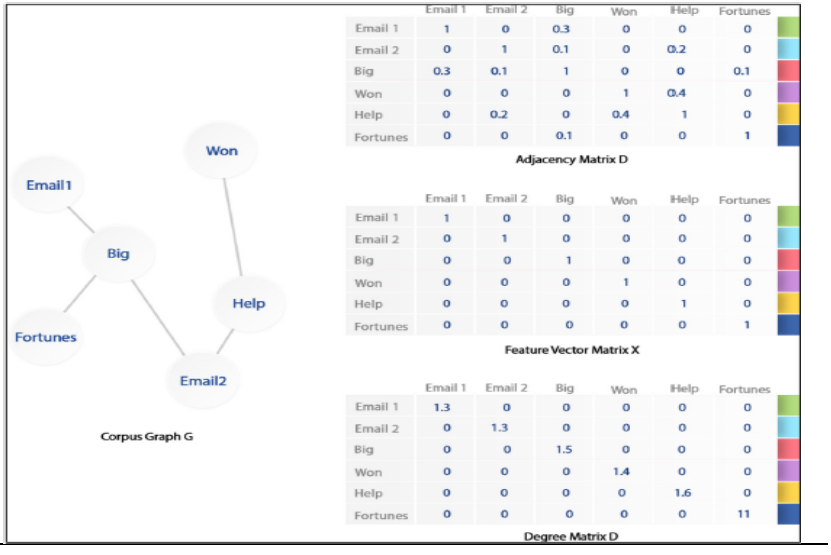
4.2. مفهوم تمثيل وإنشاء الرسم البياني للشبكة العصبية متعددة الطبقات (MLP) :

بعد تنظيف مجموعة البيانات، فإن الخطوة التالية تتمثل في إنشاء رسم بياني كبير واحد لمجموعة البريد الإلكتروني بالكامل، باستخدام الكلمات ورسائل البريد الإلكتروني كعُقد "nodes". تعتمد الحواف "edges" التي تربط عُقد الكلمة على معلومات التواجد المشترك بين الكلمتين. يتم إنشاء الحواف "edges" بين الكلمة والبريد الإلكتروني باستخدام تردد الكلمة "word frequency" وتردد رسالة البريد الإلكتروني للكلمة " word's email

http://www.doi.org/10.62341/mksk1904

frequency". ومن ثم تصبح وتتحول مشكلة تصنيف النص إلى مشكلة تصنيف العُقدة "node".

فإننا نشير إلى الرسم البياني بـ "G"، حيث "G" هو رسم بياني غير موجه و "$G = (V, E)$"، حيث "V" ترمز لمجموعة العُقد "N"، و"E" ترمز لمجموعة الحواف "Edges". العُقد "N" تمثل الكلمات ورسائل البريد الإلكتروني. لإنشاء الرسم البياني، يتم تقديم المصفوفة المتجاورة "A" حيث $A\hat{I}R^{NxN}$ ، في المصفوفة "A"، يتم تمثيل كل كلمة وكل رسالة في مجموعة جسم البريد الإلكتروني بالكامل كصفوف وأعمدة لتسجيل الارتباط بينهما. القيم الموجودة في المصفوفة "A" هي عبارة عن أوزان "weights" العلاقات بين العُقد "nodes". يتم إنشاء مصفوفة قطرية يُشار إليها بالرمز "D" من مصفوفة الدرجة "A"، وتحتوي على قيم الارتباط بين كل عُقدة والعُقد الأخرى، $D\hat{I}R^{NxN}$ و $D_{ii} = \sum_i A_{ij}$. يتم تمثيل مصفوفة الهوية الخاصة بـ "A" بالمتجه "X"، حيث "X" عبارة عن مصفوفة متجه للميزات، والتي تُظهر كل رسالة بريد إلكتروني والكلمات المرتبطة بها في متجه واحد. باستخدام تمثيل الرسم البياني لـ "G"، يوضح "الشكل 6–" مفهوم إنشاء رسم بياني للشبكة العصبية متعددة الطبقات (MLP).



الشكل 6: مفهوم تمثيل الرسم البياني للشبكة العصبية متعددة الطبقات (MLP) .

تمثل A_{IJ} علاقة الحافة بين العُقدتين "i" و"j"، ويتم حساب قيمة الحواف على النحو التالي: إذا كانت "i" و"j" هما نفس العقدة أو كلمتان متطابقتن، فإن القيمة تساوي 1؛ وإذا

http://www.doi.org/10.62341/mksk1904

كانت "i" و"j" كلمتان مختلفتان، فإن القيمة هي الرابط أو العامل المشترك بينهما. يستخدم مصطلح التردد "TF"– ومصطلح تردد المستند العكسي "IDF" لحساب الحافة بين الكلمة التي يتم الإشارة إليها باستخدام "i" والبريد الإلكتروني المشار إليه بواسطة "j".

يتم استخدام مصطلح "TF" و"IDF" بشكل شائع في تطبيقات البحث عن المستندات أو الوثائق؛ وإنه يمثل علاقة الارتباط بين الكلمة وبريد إلكتروني محدد. يتم حساب قيمة "TF" و"IDF" بقسمة عدد المرات التي تظهر فيها الكلمة في رسالة بريد إلكتروني على العدد الإجمالي لجميع رسائل البريد الإلكتروني التي تحتوي على هذه الكلمة. لذلك، كلما زاد عدد رسائل البريد الإلكتروني التي تحتوي على هذه الكلمة، انخفضت أو قلت قيمة "TF" –"IDF" للكلمة لأنه لا يوجد ارتباط كبير بين البريد الإلكتروني والكلمة، كما يظهر في العديد من رسائل البريد الإلكتروني الأخرى.

يتم استخدام المعلومات المتبادلة النقطية "PMI"، والمعروفة أيضًا باسم الكلمات المتزامنة أو "التكرار المشترك للكلمة". يقوم "PMI" بحساب ارتباطات الكلمات كما هو موضح في المعادلة التالية:

$$\mathrm{PMI}(\mathrm{I},\mathrm{J}) = \frac{\mathrm{P(I,J)}}{\mathrm{P(I)\ P(J)}} \quad \ldots\ldots (1)$$

حيث $P(I,J)$ هو احتمال تقارب أوتجاور الكلمة "I" والكلمة "J"، $P(I)$ و$P(J)$ هي احتمالية الكلمة "I" والكلمة "J" في جسم الرسالة على التوالي .

4.3. مصنف الشبكة العصبية متعددة الطبقات (MLP) :

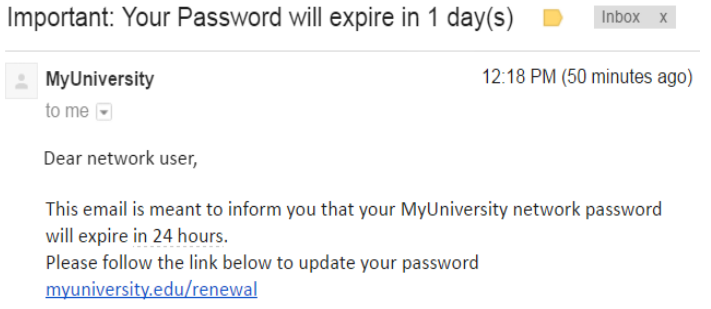
كل عُقدة "node" في الرسم البياني تأخذ معلومات جميع جيرانها من العُقد "nodes" المجاورة، بما في ذلك معلومات العقدة نفسها. ثم يتم إرسال القيمة المحسوبة إلى الشبكة العصبية "NN". قبل تمرير معلومات الجيران إلى عقدة معينة، يتم قياس متجهات العقدة أو ضبطها. لتحقيق وإجراء القياس، يتم ضرب المصفوفة "A" في مصفوفة الدرجة الخاصة بها "D". السبب وراء تحجيم وقياس المتجهات لأن العقد التي لديها العديد من الجيران لها متجهات مجمعة كبيرة لـ "v"، في حين أن العقد مع عدد قليل من الجيران لها متجهات مجمعة صغيرة، والتي يمكن أن تؤدي إلى اختفاء وتلاشى التدرجات ويمنع الشبكة العصبية "NN" من تلقي المزيد من التدريب. وبالتالي، نقوم بتوسيع نطاق هذه المتجهات

http://www.doi.org/10.62341/mksk1904

للتغلب على المشكلات المحتملة. يتم ضرب معكوس المصفوفة "D" مرتين في المصفوفة "A" لقياس كل من الصفوف والأعمدة، وتحتوي على معلومات حول جميع العُقد المجاورة، بما في ذلك العقدة نفسها. وبالتالي، سيتم قياس المصفوفة "A" لتكون ← "Ã" التي تسمى المصفوفة المجاورة :

$$\tilde{A} = D^{-0.5} A D^{-0.5} \quad \ldots\ldots (2)$$

في خوارزمية الشبكة العصبية متعددة الطبقات (MLP)، عندما نستخدم طبقة شبكة تلافيفية واحدة، يمكن للعقدة فقط جمع والتقاط المعلومات من أقرب جيرانها. لذلك، كلما تم استخدام المزيد من الطبقات التلافيفية، تم دمج المزيد من المعلومات من الجيران "nodes". وبالتالي، كلما كانت هناك حاجة إلى مزيد من المعلومات حول الجيران البعيدين (العُقد البعيدة)، يجب أن يكون لدينا المزيد من الطبقات. مهمة جمع المعلومات حول كل عقدة من جيرانها يتم تنفيذها في وقت واحد وبالتوازي مع كل عقدة. نكرر مهمة جمع المعلومات لأقرب العقد فقط باستخدام طبقتين. نحتاج عادةً إلى طبقتين أو ثلاث طبقات في الرسم البياني.


Important: Your Password will expire in 1 day(s) Inbox x

MyUniversity 12:18 PM (50 minutes ago)

to me

Dear network user,

This email is meant to inform you that your MyUniversity network password will expire in 24 hours.
Please follow the link below to update your password
myuniversity.edu/renewal

الشكل 7: مثال على رسالة تصيد عبر البريد الإلكتروني.

يتم حساب مصفوفة السمة / الميزة للطبقة الأولى على النحو التالي:

$$L^{(1)} = P\ (\tilde{A}\ X\ W_0) \quad \ldots\ldots (3)$$

حيث "Ã" هي لقياس المصفوفة المجاورة التي تم توضيح طريقة حسابها مسبقًا في المعادلة (2) على أنها:

http://www.doi.org/10.62341/mksk1904

$$\tilde{A} = D^{-0.5} AD^{-0.5} \quad \ldots\ldots (4)$$

حيث "X" هي مصفوفة الهوية "Identity matrix" لـ "A" ، و"W_0" هي مصفوفة الوزن "Weight matrix" التي يمكن حسابها باستخدام النسبة المتدرجة، و"P" هي دالة التنشيط مثل: (دالة التنشيط السينية "sigmoid activation function"، أو دالة تنشيط "ReLU activation function").

يمكن تحديد وحساب مصفوفة السمات / الميزات لبقية الطبقات على النحو التالي:

$$L^{(i+1)} = P\ (\tilde{A}\ L^{l}\ W_0) \quad \ldots\ldots (5)$$

يتم توفير مصفوفة الميزات / السمات للبريد الإلكتروني، والتي تتكون من طبقة إدخال – وطبقات مخفية – وطبقة إخراج ، كمدخلات إلى الشبكة العصبية "NN". في هذه الدراسة، تقبل طبقة الإدخال– ميزات رسائل البريد الإلكتروني كبيانات خارجية للنموذج أو المصنف، وتتعلم الطبقات المخفية – في نفس الوقت جميع عمليات الفحص في مجموعة التدريب لإنتاج وإنشاء المخرجات. توفر طبقة الإخراج – التصنيف "1" للبريد الإلكتروني المخادع (للتصيد الاحتيالي)؛ أو التصنيف "0" للبريد الإلكتروني الشرعي.

5. الأدوات، النتائج ، والمناقشات :

5.1. أدوات التجربة :

المعمل المستخدم في هذه التجربة عبارة عن جهاز حاسوب شخصي بالمواصفات التالية: معالج "Intel Core i7 processer" رباعي النواة بسرعة تردد 2.3 جيجا هرتز "GHz"، وذاكرة وصول عشوائي "LPDDR4X RAM" بسعة 32 جيجابايت "GB" و3733 ميجاهرتز "MHz"، ونظام تشغيل ويندوز "Windows" 64 بت "bit". بالإضافة إلى ذلك ، تم استخدام لغة البرمجة بايثون "Python language".

5.2. مجموعة البيانات المستخدمة في الاختبارات :

في هذا المشروع البحثي، استخدمنا مجموعة البيانات الأولية والاحتيالية ("Collection of Fraud email [21]")، وهي بيانات متاحة للعامة، والتي تُستخدم في تجارب تصنيف البريد الإلكتروني. مجموعة البيانات متوازنة ومُصنَّفة لاستخدامها في التعلم الآلي (ML)

http://www.doi.org/10.62341/mksk1904

الخاضع للإشراف. تتم تسمية وتصنيف رسائل التصيد الاحتيالي ورسائل البريد الإلكتروني الشرعية بالرقم (1) و(0)، على التوالي. يحتوي السجل "record" على نص رسائل البريد الإلكتروني. تتكون مجموعة البيانات المستخدمة مما يلي:

- 3685 رسالة بريد إلكتروني للتصيد الاحتيالي "phishing Emails"، "الشكل-7" الموضح أعلاه هو مثال على رسالة بريد إلكتروني للتصيد الاحتيالي.
- 4894 رسالة بريد إلكتروني شرعي "legitimate Emails".

5.3. إعداد / ضبط المُصنِّف "MLP" :

تحتوي خوارزميات الشبكات العصبية لـ "MLP" على معاملات "parameters" مختلفة لتعيينها، مثل: (عدد طبقات مصنف الشبكة العصبية (MLP)، الحد الأقصى لعدد التكرارات ، والإيقاف المبكر لحجم نافذة نص رسالة البريد الإلكتروني، ... إلخ). وفقًا لعمل هذا المشروع البحثي، تم تعيين وضبط معاملات "parameters" لتحقيق أقصى وأعلى درجة من الدقة، مثل ليكون: (معدل التعلم "learning rate": 0.1؛ تعديل الخلط "Shuffle" = "True"؛ التوقف المبكر: بعد 10 فترات؛ عدد طبقات "MLP": طبقتان أو ثلاث طبقات؛ دالة الفاقد / الخسارة "Loss function": خطأ في معدل نقل المعلومات برسالة معينة عبر "cross-entropy error").

5.4. التقييم والنتائج :

لتقييم فعالية خوارزمية مصنف التصيد الاحتيالي لهذا المشروع البحثي، تم تقسيم مجموعة البيانات الأولية، وهي مجموعة من الحالات والبيانات المحددة مسبقًا، إلى مجموعتين منفصلتين: مجموعة التدريب ومجموعة الاختبار. أولاً، تم استخدام مجموعة التدريب- للسماح بتدريب الشبكة العصبية للخوارزمية الهجينة، التي تجمع بين أساليب خوارزميات "MLP" وتقنيات معالجة اللغة الطبيعية "NLP"، لمعرفة كيفية تصنيف رسائل البريد الإلكتروني. ثانيًا، احتفظت واحتوت مجموعة الاختبار-على البيانات المطلوبة التي يتعين تغذيتها للشبكات العصبية "NN" للخوارزمية الهجينة، دون تصنيفها (مجموعة البيانات غير المسماة).

بعد ذلك، يتم استخدام الشبكة العصبية "NN" للمُصنِّف، للتنبؤ بفئات تصنيف رسائل البريد الإلكتروني. ثم يتم استخدام مجموعة الاختبار- لتقييم أداء تنبؤات الشبكة العصبية

http://www.doi.org/10.62341/mksk1904

من خلال مقارنة التنبؤات بالقرارات المصنفة والتي تم تحديدها يدويًا. باتباع مقاييس التقييم المستخدمة في الدراسات البحثية [10؛ 22؛ 23؛ 24]، استخدمنا مقاييس: (الانضباط، الدقة، والاسترجاع) لتقييم نموذج المُصنِّف المقترح في هذا البحث ومقارنتها بأحدث نتائج النماذج الأخرى المنشورة.

المعدل الإيجابي الزائف / الكاذب "False Positive (FP)" هو عدد رسائل البريد الإلكتروني المشروعة "legitimate (ham)" التي تم تصنيفها بشكل خاطئ على أنها تصيد احتيالي، المعدل السلبي الحقيقي / الصحيح "True Negative (TN)" هو عدد رسائل البريد الإلكتروني المشروعة المصنفة على أنها شرعية، المعدل السلبي الخاطئ / الكاذب "False Negative (FN)" هو عدد رسائل البريد الإلكتروني المخادعة "phishing emails" التي تم تصنيفها بشكل خاطئ على أنها شرعية، والمعدل الإيجابي الحقيقي / الصحيح "True Positive (TP)" هو عدد رسائل التصيد الاحتيالي الإلكترونية المصنفة كرسائل بريد إلكتروني للتصيد الاحتيالي. يعرض "الجدول 1" مصفوفة التصنيف الموضحة أدناه في الجدول.

الجدول 1 – مصفوفة التصنيف.

	Predict	
Actual	0 - Legitimate (ham)	1 - phishing
0 - Legitimate (ham)	TN (True Negative)	FP (False Positive)
1 - phishing	FN (False Negative)	TP (True Positive)

لتقييم النماذج، يتم استخدام المقاييس التالية:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad \ldots(6)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad \ldots(7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad \ldots(8)$$

$$\text{FPR} = \frac{FP}{FP + TN} \quad \ldots(9)$$

http://www.doi.org/10.62341/mksk1904

هناك طريقتان تستخدمان بشكل شائع لتقييم فعالية المصنف: (طريقة "Holdout split"، وطريقة "k-fold cross-validation"[25]). في طريقة "Holdout split"، يتم تقييم أداء المصنف عن طريق استخدام مجموعة الاختبار بدلاً من مجموعة التدريب لحساب أو قياس معدل الخطأ في مجموعة البيانات الجديدة. من الناحية العملية، يتم تقسيم مجموعة البيانات إلى جزأين: (جزء من ثلثي مجموعة البيانات للتدريب، والثلث المتبقي من مجموعة البيانات للاختبار). في طريقة "k-fold cross-validation"، يتم تقسيم مجموعة التدريب بشكل عشوائي إلى مجموعات منفصلة (fold-1st , fold-2nd , fold-3rd , …. , fold-K) ذات حجم متساوي، حيث يكون لكل جزء منها نفس التوزيع لحجم الفئة تقريبًا. يتم تدريب المصنف عدد "k" من المرات، باستخدام مجموعة اختبار مختلفة في كل مرة؛ حيث يكون الخطأ المقدّر هو متوسط عدد الأخطاء.

بالنسبة لتقييم الخوارزمية الهجينة لنموذج المُصنِّف المقترح لهذه الورقة البحثية، تم تطبيق واستخدام كلتا الطريقتين. تعتمد طريقة الـ "Holdout split" على تقسيم البيانات إلى مجموعتين من البيانات، واحدة للتدريب والأخرى للاختبار. استخدمنا 70٪ من البيانات (6005 رسالة) للتدريب، بينما 30٪ (2574 رسالة) للاختبار. أضف إلى ذلك، تم استخدام طريقة "**3**-fold cross-validation"، من خلال تقسيم مجموعة البيانات إلى ثلاث مجموعات؛ ويتم إجراء تدريب المُصنِّف المقترح ثلاث مرات، في كل مرة بمجموعة مختلفة يتم وضعها كمجموعة اختبار. يبين "الجدول 2" نتائج التحقق الواردة أدناه في الجدول.

الجدول 2 – نتائج التحقق.

Fold	iteration	Precision	Recall	F1-score	Avg precision	Avg recall	Avg f1-score	Accuracy
1st Fold	0	0.982	0.991	0.986	0.9867	0.9868	0.9867	0.9868
	1	0.991	0.982	0.986				
2nd Fold	0	0.990	0.988	0.989	0.9911	0.9900	0.9909	0.9913
	1	0.992	0.993	0.9934				
3rd Fold	0	0.969	0.992	0.980	0.9820	0.9855	0.9836	0.9841
	1	0.994	0.9781	0.986				

http://www.doi.org/10.62341/mksk1904

جدول مصفوفة التصنيف، والذي تمت مناقشته سابقاً يتم عرض نتائجه في "الجدول 3".

الجدول 3 – نتائج مصفوفة التصنيف.

Actual	Predict 0 – Legitimate (ham)	1 – phishing
0 – Legitimate (ham)	4796	98
1 – phishing	74	3611

لتقييم فعالية المصنف، من المتوقع أن يكون معدل قيم "TP" و"TN" كبير، في حين أن معدل قيم "FP" و"FN" صغير؛ تؤكد النتائج وتدعم ما هو متوقع ويمكن استخدامها للحصول على المزيد من مقاييس التقييم. عند تطبيق القيم لـ "FP" و"TN" في المعادلة رقم: (9)، فأن معدل "FPR" المُقاس هو 0.015، مما يشير إلى احتمال ضئيل لتصنيف رسائل البريد الإلكتروني الشرعية على أنها تصيد احتيالي. يتم حساب تقييم المصنف باستخدام (معدل الدقة الشاملة "Accuracy"، معدل الانضباط للدقة "Precision"، ومقياس الاسترجاع "Recall") كما هو موضح في هذا القسم 5.4 في المعادلات رقم: (6)، (7)، و(8). يعرض "الجدول 4" قيم نتائج مقاييس التقييم الموضحة أدناه في الجدول.

الجدول 4 – نتائج مقاييس التقييم.

Accuracy	98.2%
Precision	98.5%
Recall	98.3%
f–measure	98.5%

يتم حساب مقياس "f-measure" باستخدام المتوسط التوافقي للانضباط "Precision" والاسترجاع "Recall"، كما هو موضح في المعادلة رقم: (10).

$$f - measure = \frac{(Precision + Recall)}{2} \quad \ldots(10)$$

http://www.doi.org/10.62341/mksk1904

تتم مقارنة مقاييس النتائج الخاصة بمُصنِّفنا المقترح: (مقياس الاسترجاع "Recall "، معدل الدقة الشاملة "Accuracy "، معدل الانضباط للدقة "Precision "، ومقياس المتوسط التوافقي "f-measure") مع أحدث نتائج الدراسات الأخرى المنشورة، لتقييم دقة وفعالية نموذجنا للخوارزمية الهجينة. يقدم "الجدول 5" القيم المقارنة لمقاييس نتائج الأعمال ذات الصلة من حيث ("Recall "، "Accuracy "، "Precision "، و"f-measure")، والمعروضة أدناه في الجدول.

الجدول 5 - مقارنة الأعمال ذات الصلة من حيث معدل الدقة الشاملة، معدل الانضباط للدقة، مقياس الاستدعاء ومقياس المتوسط التوافقي.

Reference	Technique	Accuracy	Precision	Recall	f-measure
Our model. **(نموذجنا)**	MultiLayer Perceptron (MLP) Neural Network	98.2%	98.5%	98.3%	98.55%
(Lai S ,et al., 2015). [27]	Text classification for phishing detection based on RCNN.	96.94%	_	_	_
(Nguyen M ,Nguyen T ,Nguyen TH., 2018a). [19]	Deep learning hierarchical long short-term memory networks (H-LSTMs).	98.37%	97%	95%	96%
(HalgašL, AgrafiotisI, NurseJRC., 2020). [2]	Deep learning	96.74%	97.45%	95.98%	96.71%
(Peng T, Harris I, Sawa Y., 2018a). [4]	Machine learning algorithm and NLP	_	95%	91%	_
(Bergholz A ,et al ., 2008). [26]	Machine learning algorithm using semantic features	98.48%	97.95%	97.93%	97.94%
(Fang Y, et al., 2019). [5]	Deep learning and NLP	98.39%	97.7%	97.6%	97.65%

يُظهر "الجدول 5"، أن مُصنِّف "MLP" يتنافس مع مصنِّفات التعلم الآلي (ML) الأخرى في اكتشاف البريد الإلكتروني للتصيد الاحتيالي بمعدل دقة مرتفع. من الجيد ملاحظة أن التقنية المستخدمة في دراسة بيرغولز وآخرون "Bergholz A,et al., [26]" التي حققت دقة أعلى، استخدمت تقنيات التعلم الآلي (ML) اعتمادًا على عمليات استخراج الميزات، والتي تتطلب وتستلزم إشراك خبير المجال للتدخل في العملية من أجل تقليل تعقيد

http://www.doi.org/10.62341/mksk1904

البيانات وجعل الأنماط أكثر وضوحًا لخوارزميات التعلم. ومع ذلك، فإن نموذجنا للمُصنِّف المقترح يلغي الحاجة إلى تدخل خبير المجال واستخراج الميزات المصنوعة يدويًا. علاوة على ذلك، في دراسة بيرغولز وآخرون "Bergholz A,et al., [26]"، فقد استخدموا ثماني وظائف لميزات الارتباط غير متوافقة أو لن تعمل مع رسائل البريد الإلكتروني التي لا تحتوي على روابط "links". بالإضافة إلى ذلك ، في دراسة فانغ وآخرون "Fang Y, et al., [5]" ، فقد استخدموا كلاً من عنوان وتوصيف البريد الإلكتروني لتحديد واكتشاف الرسائل المخادعة، بينما يقوم نموذجنا للمُصنِّف المقترح بتحليل محتوى النص الأساسي لجسم رسالة البريد الإلكتروني، مما يقلل من تعقيد الذاكرة "Memory Complexity". إلى جانب ذلك، يعمل نموذج المُصنِّف المقترح بشكل أفضل مع معدل الانضباط للدقة "Precision" ومقياس الاسترجاع "Recall" مقارنةً بدراسة نغوين وآخرون "Nguyen M,et al.,[19]" ، مما يؤدي إلى تقليل وقت التعلم والتصنيف.

بالمقارنة مع طرق التعلم العميق (DL) الأخرى المستخدمة لاكتشاف رسائل البريد الإلكتروني المخادعة التي تتضمن صورًا، فإن هذا النموذج أكثر فاعلية لأنه يتضمين معالجة الميزات القائمة على الصور، مما يؤدي إلى زيادة الكفاءة في وقت التشغيل وتقليل المساحة. علاوة على ذلك، وجدنا أيضًا أنه حتى في رسائل البريد الإلكتروني التي تحتوي على صور، فإن نموذجنا والذي يحلل محتوى نص الرسالة والمعلومات المتعلقة بالروابط "links"، كان أكثر فاعلية وملاءمة في اكتشاف التصيد الاحتيالي.

من الناحية النظرية، تحتاج ذاكرة خوارزمية "MLP" إلى وقت التشغيل أو زمن تنفيذ وقدرة = "O(E)"، وتعقيدها الزمني "Time Complexity" وقدرة = "O(LED + LND2)"؛ حيث تشير ("L" إلى عدد الطبقات "layers" ، "N" إلى العدد الإجمالي للعقد "nodes"، "D" إلى بُعد الميزة بالنسبة للعقد، و"E" إلى عدد حواف الرسم البياني للعقد). وفقًا للتجارب في هذا المشروع البحثي، فقد أظهرت أن وقت تشغيل أو زمن تنفيذ خوارزمية النموذج: يبلغ في المتوسط "3.89" دقيقة–لتدريب 6005 رسالة بريد إلكتروني، ولاختبار 2574 رسالة بريد إلكتروني يستغرق المتوسط "1.08" ثانية؛ والذي يعتبر وقتًا قصيرًا مقارنة بدراسة نغوين وآخرون " Nguyen M,et al.,[19]"، والذي استغرق من "61" إلى "181" دقيقة لأقل من 6000 رسالة بريد إلكتروني.

http://www.doi.org/10.62341/mksk1904

6. الاستنتاجات والعمل المستقبلي :

تُعد رسائل البريد الإلكتروني للتصيد الاحتيالي أحد أسرع الجرائم الإلكترونية انتشارًا على الإنترنت، حيث تؤثر وتلحق الضرر بالشركات والأفراد على حدٍ سواء، مما يؤدي إلى خسائر تقدر بمليارات الدولارات سنويًا. تتغير أساليب وتقنيات التصيد الاحتيالي بسرعة وتعتمد على ذكاء المخترق / الهاكر "hacker". لذلك، فإن الاستفادة من التعلم الآلي العميق (DL) للتنافس مع الذكاء البشري ومطابقته هو الحل الأمثل.

الهدف الرئيسي من هذه الدراسة هو اقتراح وتطوير نموذج فعال قادر على تحسين أداء تصنيف رسائل البريد الإلكتروني واكتشافها على أنها شرعية أو احتيالية بدقة عالية. قمنا بتطبيق نموذج تصنيف اكتشاف رسائل البريد الإلكتروني للتصيد الاحتيالي باستخدام معالجة اللغة الطبيعية (NLP) وخوارزمية التعلم العميق القائمة على الشبكة العصبية متعددة الطبقات (MLP). تم إختبار النموذج باستخدام منهجية تعلم خاضعة للإشراف، وقد ثبت أن المصنِّف المقترح يعمل بشكل يحسن دقة التصنيف والأداء مقارنة بالطرق الأخرى؛ وأنجز النموذج: (تنفيذ التدريب وإتمامه في وقت قصير، ومعدل إيجابي حقيقي (TP) مرتفع، معدل سلبي حقيقي (TN) مرتفع، مقياس استرجاع "Recall " عالي، معدل الانضباط للدقة "Precision " عالي، مقياس المتوسط التوافقي "f-measure" مرتفع، ومعدل الدقة الشاملة "Accuracy " عالي).

يسلط نموذج المصنِّف المقترح الضوء على أهمية فحص الميزات النصية في جسم رسالة البريد الإلكتروني لاكتشاف التصيد الاحتيالي، لأن الميزات النصية هي اتجاه بحثي جديد نسبيًا في مجال اكتشاف التصيد عبر البريد الإلكتروني؛ وهناك أيضًا عدد قليل جدًا من الدراسات التي ركزت على جسم البريد الإلكتروني. هذه الدراسة هي واحدة من أولى الدراسات، على حد علمنا، التي بحثت وفحصت في كيفية استخدام خوارزمية هجينة تجمع بين أساليب الـ("MLP" & "NLP") لاكتشاف وتحديد التصيد الاحتيالي في جسم النص الأساسي لرسالة البريد الإلكتروني. وكذلك لتسليط الضوء على تبني المزيد من الأبحاث في مجال كيفية التحقق والكشف عن التصيد الاحتيالي في النص، سواء في رسائل البريد الإلكتروني أو رسائل وسائل التواصل الاجتماعي أو حتى رسائل التصيد لتطبيقات أخرى. علاوة على ذلك، يمكن للمُصنِّف المقترح أن يكتشف وبدقة رسائل البريد

http://www.doi.org/10.62341/mksk1904

الإلكتروني التصيدية التي لم يسبق رؤيتها من قبل، استنادًا إلى النص الأساسي؛ لذلك، نعتقد أنه فعال في اكتشاف هجمات التصيد الاحتيالي من المرة الأولى (في اليوم صفر).

بالإضافة إلى ذلك، نعتزم ونخطط لإجراء دراسة مستقبلية باستخدام مجموعة رسائل التواصل الاجتماعي ورسائل التطبيقات الأخرى. في بحثنا المستقبلي، نظرًا لأن مفهوم النموذج المقترح يعتمد على تصنيف النص لجسم الرسالة، فسيتم إجراء المزيد من التحقيقات لتقييم فعاليته في مهام تصنيف واكتشاف الرسائل الاحتيالية المكتوبة بواسطة تطبيقات أخرى.

7. الخـــلاصـــة والـتوصـيات:

- فاعلية النموذج: **لقد** أثبت النموذج المقترح كفاءته في اكتشاف رسائل البريد الإلكتروني الاحتيالية، حيث حقق النموذج المقترح أداءً عاليًا جدًا في تصنيف الرسائل، بدقة تجاوزت 98%.
- سرعة في التصنيف: **بالإضافة** إلى الدقة العالية، كان النموذج سريعاً في تصنيف الرسائل.
- أهمية التقنيات المستخدمة لاكتشاف التصيد الاحتيالي: **أكدت** الدراسة أهمية الاستفادة من دمج تقنيات معالجة اللغة الطبيعية والتعلم العميق في هذا المجال.
- فتح آفاق جديدة: يفتح هذا البحث آفاقاً جديدة لتطوير أنظمة أمن المعلومات من خلال الجمع بين معالجة اللغة الطبيعية (NLP) وتقنيات التعلم العميق القائمة على الشبكة العصبية متعددة الطبقات (MLP) لتحليل محتوى الرسائل واكتشاف الدلائل التي تشير إلى ما إذا كانت احتيالية أم شرعية.
- تطوير النموذج: يقدم هذا البحث حلاً واعداً يفتح الباب أمام تطوير أنظمة أمن المعلومات، ويمكن توسيع نطاق البيانات ليشمل مجموعات بيانات أكبر وأكثر تنوعًا لتحليل وفحص ميزات إضافية مثل رسائل وسائل التواصل الاجتماعي أو حتى رسائل التصيد للتطبيقات الأخرى.
- تطبيق النموذج في أنظمة البريد الإلكتروني: يمكن دمج النموذج في أنظمة البريد الإلكتروني لحماية المستخدمين من الرسائل الاحتيالية.

http://www.doi.org/10.62341/mksk1904

References :

[1]- Akanbi , O. A., Amiri, I. S., & Fazeldehkordi , E. (2015). A machine-learning approach to phishing detection and Defense. https://doi.org/10.1016/c2014-0-03762-8 .

[2]- Halgaš, L., Agrafiotis, I., & Nurse, J. R. (2020). Catching the phish: Detecting phishing attacks using recurrent neural networks (rnns). Information Security Applications, 219–233. https://doi.org/10.1007/978-3-030-39303-8_17 .

[3]- Khonji, M., Iraqi, Y., & Jones, A. (2013). Phishing detection: A literature survey. IEEE Communications Surveys & Tutorials, 15(4), 2091–2121. https://doi.org/10.1109/surv.2013.032213.00009.

[4]- Peng, T., Harris, I., & Sawa, Y. (2018**a**). Detecting phishing attacks using natural language processing and machine learning. 2018 IEEE 12th International Conference on Semantic Computing (ICSC). https://doi.org/10.1109/icsc.2018.00056 .

[5]- 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. https://doi.org/10.1109/access.2019.2913705 .

[6]- McAfee 2020 The hidden cost of cybercrime by McAfee. McAfee Blog. (2020, December 21). Retrieved December 24, 2022, from https://www.mcafee.com/blogs/other-blogs/executive-perspectives/the-hidden-costs-of-cybercrime-on-government/ .

[7]- Sadique, F., Kaul, R., Badsha, S., & Sengupta, S. (2020). An automated framework for real-time phishing URL detection. 2020 10th Annual Computing and Communication Workshop and Conference (CCWC). https://doi.org/10.1109/ccwc47524.2020.9031269 .

[8]- Rosenthal, M. (2022, May 13). Phishing statistics (updated 2022) - 50+ important phishing stats. Tessian. Retrieved December 24, 2022, from https://www.tessian.com/blog/phishing-statistics-2020/.

[9]- Verizon- DBIR 2020 Data Breach Investigations Report - Verizon. (n.d.). Retrieved December 19, 2022, from https://www.verizon.com/business/en-gb/resources/reports/2020-data-breach-investigations-report.pdf .

http://www.doi.org/10.62341/mksk1904

[10]- Gupta, B. B., Arachchilage, N. A., & Psannis, K. E. (2018). Defending against phishing attacks: Taxonomy of methods, current issues and future directions. Telecommunication Systems, 67(2), 247–267. https://doi.org/10.1007/s11235-017-0334-z .

[11]- APWG . Anti-Phishing Working Group 2020 Report: A study of the scope and distribution of phishing. APWG. Retrieved December 19, 2022, from https://apwg.org/phishing-landscape-2020-a-study-of-the-scope-and-distribution-of-phishing/ .

[12]- Microsoft Digital Defense Report 2020: Cyber threat sophistication on the rise. Microsoft Security Blog. Retrieved December 24, 2022, from https://www.microsoft.com/en-us/security/blog/2020/09/29/microsoft-digital-defense-report-2020-cyber-threat-sophistication-rise/

[13]- Benenson, Z., Gassmann, F., & Landwirth, R. (2017). Unpacking spear phishing susceptibility. Financial Cryptography and Data Security, 610–627. https://doi.org/10.1007/978-3-319-70278-0_39

[14]- Yao, L., Mao, C., & Luo, Y. (2019). Feature selection based on term frequency for text classification using Multilayer Perceptron (MLP) . Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 7370–7377. https://doi.org/10.1609/aaai.v33i01.33017370

[15]- Al-Qerem, A., Alauthman, M., Almomani, A., & Gupta, B. B. (2019). IOT transaction processing through cooperative concurrency control on FOG–Cloud Computing Environment. Soft Computing, 24(8), 5695–5711. https://doi.org/10.1007/s00500-019-04220-y

[16]- Alieyan, K., Almomani, A., Anbar, M., Alauthman, M., Abdullah, R., & Gupta, B. B. (2019). DNS rule-based schema to botnet detection. Enterprise Information Systems, 15(4), 545–564. https://doi.org/10.1080/17517575.2019.1644673

[17]- Jain, A. K., & Gupta, B. B. (2019). A machine learning based approach for phishing detection using hyperlinks information. Journal of Ambient Intelligence and Humanized Computing, 10(5), 2015–2028. https://doi.org/10.1007/s12652-018-0798-z

[18]- Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in Machine Learning: A new perspective. Neurocomputing, 300, 70–79 Elsevier . https://doi.org/10.1016/j.neucom.2017.11.077

[19]- Nguyen , M., Nguyen, T., & Nguyen , T. H. (2018**a**). A deep learning model with hierarchical LSTMs and supervised

http://www.doi.org/10.62341/mksk1904

attention for anti-phishing. 2018 10th Asian Control Conference (ASCC). https://doi.org/10.1109/ascc.2015.7244834

[20]- Deng, L. (2014). Deep learning: Methods and applications. Foundations and Trends® in Signal Processing, 7(3-4), 197–387. https://doi.org/10.1561/2000000039

[21]- Radev, D. (2008, August 2). Fraud email dataset-CLAIR collection of fraud email, ACL Data and Code Repository. Kaggle. Retrieved December 24, 2022, from https://www.kaggle.com/datasets/llabhishekll/fraud-email-dataset

[22]- Nguyen , M., Nguyen, T., & Nguyen , T. H. (2018**b**). Phishing identification: An efficient neuro-fuzzy model without using rule sets. 2018 10th Asian Control Conference (ASCC). https://doi.org/10.1109/ascc.2015.7244631

[23]- Peng, T., Harris, I., & Sawa, Y. (2018**b**). Detecting phishing attacks using natural language processing and machine learning. 2018 IEEE 12th International Conference on Semantic Computing (ICSC). https://doi.org/20.3405/icsc.2018.00073

[24]- IC3 releases 2020 internet crime report. FBI. Retrieved December 24, 2022, from https: //www.ic3.gov/Media/PDF/AnnualReport/2020 _ IC3Report.pdf

[25]- Rithchie, N. (2020). Evaluating model estimation processes for diagnostic classification models. https://doi.org/10.31237/osf.io/vke5u

[26]- Bergholz, A., Let, Š., Šedivý, J., & Hlaváč, V. (2008). Improved phishing detection using model-based features. Proceedings of the 5th International Conference on Information Systems Security and Privacy. https://doi.org/10.5220/0007314202520256

[27]- Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent convolutional neural networks for text classification. Proceedings of the AAAI Conference on Artificial Intelligence, 29(1). https://doi.org/10.1609/aaai.v29i1.9513.