South American Publishing

SAP Gamification and Augmented Reality. 2025; 3:91
doi: 10.56294/gr202591

ORIGINAL

Guardians of the Web: Harnessing Machine Learning to Combat Phishing Attacks

Guardianes de la Web: Aprovechando el Aprendizaje Automático para Combatir los Ataques de Phishing

Mowafaq Salem Alzboon[1], Mohammad Subhi Al-Batah[1], Muhyeeddin Alqaraleh[2], Faisal Alzboon[3], Lujin Alzboon[3]

[1]Jadara University, Faculty of Information Technology. Irbid, Jordan.
[2]Zarqa University, Faculty of Information Technology. Zarqa, Jordan.
[3]Caucasus International University (CIU), Dental Medicine. Tbilisi, Georgia.

Cite as
Salem Alzboon M, Subhi Al-Batah M, Alqaraleh M, Alzboon F, Alzboon L. Guardians of the Web: Harnessing Machine Learning to Combat Phishing Attacks. SAP Gamification and Augmented Reality. 2025; 3:91. https://doi.org/10.56294/gr202591

Publication History
Submitted: 19-09-2024
Revised: 25-11-2024
Accepted: 24-01-2025
Published: 25-01-2025

Corresponding Author
Mowafaq Salem Alzboon

ABSTRACT

Phishing remains one of the most dangerous threats to internet users and organizations today since it utilizes spoofed websites to coax users into revealing their data. This paper focuses on the effectiveness of algorithms in detecting such abusive websites. It goes on to analyze the dataset of phishing and non- phishing URLs providing explanatory attributes such as domain registration date, URL length or the existence of HTTPS. The models studied include Decision Tree, Random Forest, and Support Vector Machines. The results found that the Random Forest algorithm had the best performance of 97 % in terms of classification accuracy, and Support Vector Machines performed the best in terms of generalization accuracy with precision and recall values of 0,92 and 0,95, respectively. The study investigates feature selection and determinants of URL structural features which are crucial in determining the efficiency of detection. Also, to enhance model assessment the stratified 10-fold cross-validation technique was performed to reduce bias and variance. These Results show the prospect of One Layer Neural Networks as a tool to improve Phishing Detection Systems and help to provide low-cost and fast solutions for current or future cyberspace struggles. This work aims to increase confidence in online security applications against modern phishing methods. The proposed modifications will help strengthen counter measures against phishing attacks in a shifting technological context while also working towards sustaining the organizations and thus require further inquiry into the facets such as the applicability of sophisticated artificial intelligence techniques the use of useful yet diverse sets of data and the incorporation of explainable intelligent systems.

Keywords: Phishing; Website Detection; Machine Learning; Feature Extraction; Cybersecurity.

RESUMEN

El phishing sigue siendo una de las amenazas más peligrosas para los usuarios de Internet y las organizaciones hoy en día, ya que utiliza sitios web falsos para engatusar a los usuarios para que revelen sus datos. Este artículo se centra en la eficacia de los algoritmos para detectar estos sitios web abusivos. A continuación, analiza el conjunto de datos de URL de phishing y no phishing proporcionando atributos explicativos como la fecha de registro del dominio, la longitud de la URL o la existencia de HTTPS. Los modelos estudiados incluyen el árbol de decisión, el bosque aleatorio y las máquinas de vectores de apoyo. Los resultados mostraron que el algoritmo Random Forest obtuvo el mejor rendimiento en términos de precisión de clasificación, con un 97 %, y que las máquinas de vectores de apoyo obtuvieron los mejores resultados en términos de precisión de generalización, con unos valores de precisión y recuperación de 0,92 y 0,95, respectivamente. El estudio investiga la selección de características y los determinantes de las características estructurales de la URL, que son cruciales para determinar la eficacia de la detección. Además, para mejorar la evaluación del modelo se aplicó la técnica de validación cruzada estratificada de 10 veces para reducir el sesgo y la varianza. Estos resultados muestran las perspectivas de las redes neuronales de una capa como herramienta para mejorar los sistemas de detección de phishing y ayudar a proporcionar soluciones rápidas y de bajo coste para las luchas actuales o futuras en el ciberespacio. Este trabajo tiene como objetivo aumentar la confianza en las aplicaciones de seguridad en línea contra los métodos modernos de phishing. Las modificaciones propuestas ayudarán a fortalecer las contramedidas contra los ataques de phishing en un contexto tecnológico cambiante, mientras que también trabajan para sostener las organizaciones y por lo tanto requieren una mayor investigación en las facetas tales como la aplicabilidad de técnicas sofisticadas de inteligencia artificial el uso de conjuntos útiles pero diversos de datos y la incorporación de sistemas inteligentes explicables.

© 2025; The Authors. This Open access article is distributed under the Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Palabras clave: Phishing; Detección de Sitios Web; Aprendizaje Automático; Extracción de Características; Ciberseguridad.

INTRODUCTION

As the world becomes more digitalized and people become more reliant on the internet for communication and transactions, phishing has become a great threat to users, organizations, and corporates across the globe. Phishing is considered a type of cybercrime and involves the conning of users by convincing them to provide sensitive information, including usernames, passwords, and financial details. It is done by cyber criminals masquerading as trusted entities. These impersonated entities exploit the provided information for malicious and financial purposes. Considering the rapid advancement in phishing tactics – advanced email campaigns, social engineering techniques, or deceptive websites – the need for effective detection and prevention strategies is becoming apparent.[1,2]

Phishing attacks are extremely dynamic and continue to evolve, which is why conventional methods such as email filters and blacklists are not ideal or effective. Blacklists are not enough to address and eliminate the new threats that arise, and although they are easy to implement, they are not reliable in detecting newly created phishing websites. On the other hand, there has been a shift regarding cybersecurity with machine learning algorithms becoming more mainstream as they offer the ability to scout through large datasets, detect patterns, and make predictions, it allows for real time detection of phishing sites, this greatly enhances the responding ability of cyber specialists.[3,4]

The goal of this research is to empower users against scams and fortified borders against the interference of hackers through the use of machine learning to help detect phishing websites.

Research relies on Decision trees, random forest, and support vector machines algorithms to analyze a features set that include; content and structural, of a domain to enable it to differentiate between a phishing domain and a website. By conducting attribute extraction and evaluation through a machine learning model, such an entity can be preemptively flagged invasive in nature due to the algorithms ability to determine many factors out of the small access granted to it.[5,6]

Sharma et al. argues that the practical consequence of this model is found in the means of addressing the growing problem associated with phishing in a highly digitalized, connected world, this allows them or her to elevate their understanding and even adapt their approach of solving a phishing problem as a knowledge base is fundamentally constructed. This approach enhances the security sphere by allowing access to a better comprehension of the issues surrounding phishing and addresses the issue directly through the systematic and data oriented approach. Also, through an evaluation of how effective different machine learning algorithms are, he/she is able to evolve the accuracy and efficiency of the detection systems to be much more advanced.[7,8,9]

Having examined the role of machine learning in combating phishing, as well as in protecting sensitive information, this work contributes to the problem of developing the best methods of ensuring cybersecurity best practices. With the use of machine learning algorithms, there has been empowerment to help individuals and organizations avoid loss of sensitive data and online wealth. Essentially, this research not only expands on how such types of technologies work, but the new solutions or styles that can be applied to such problems further expands onto it greatly.

Related Work

The most pertinent danger in the online arena today is presented by those Fake websites that are aimed at revealing sensitive personal or financial details by tricking the useful individuals. This research focuses on the performance of distinguished machine learning algorithms using Random Forest, LightGBM, and XGBoost for the problem of fake website detection. The algorithms were trained and evaluated using a categorical database consisting of benign, defacement, phishing, and malware websites along with some metadata features. It was concluded that Random Forest achieved the best results with an accuracy rate of 97, which was better than LightGBM which achieved a rate of 96 percent and XGBoost which achieved 96,2 percent. The paper shows the great potential of ensemble learning algorithms when testing a fake website and suggests the prospects of development in order to increase their effective resistance against various computer attacks.[10]

In the past, URL structure, webpage content, and external features have often been integrated in the traditional methods of phishing detection, whereas the inclusion of machine learning for URL generation has attracted great attention as it constellates the novel URLs into phishing websites with considerable precision. Although the webpage, along with features embedded in them, can be examined, the method requires an exorbitant amount of resources which renders it ineffective in devices with moderate computational power. To ameliorate the problem, this paper suggests feature selection techniques aimed at URL characteristics to optimize detection. The methodology encompasses seven stages, encompassing data preparation, preprocessing, splitting the dataset into training and validation, feature selection, 10 fold cross validation, validation, and performance evaluation.[11,12]

The developed method was evaluated using two publicly available datasets. Certain parameters, TreeSHAP and Information Gain, were employed for feature ranking and selection in order to consider the 10, 15 and 20 features most suitable for the task. The features which were selected were embedded into three classifiers, Naïve Bayes, Random Forest and XGBoost, which were then evaluated based on factors like accuracy, precision and recall. The results derived from the study indicated that features ranked by TreeSHAP showed considerable improvement in detection accuracy. It was noted that XGBoost was best setup with first dataset with 15 features achieving an accuracy of 98,59 %. Meanwhile, Random Forest achieved a maximum accuracy of 90,21 %.[13,14,15]

With the help of the first dataset, Naïve Bayes maximally performed at a level of accuracy equal to 98,49 %. These results confirm the usefulness of the URL based feature selection techniques towards identifying more phishing websites thus rendering

more efforts that aim to improve cybersecurity effective.[16]

Phishing is still a danger to internet users today as attackers continue to devise new methods for tricking users into surrendering confidential information or installing software. Phishing sites have been identified using heuristic and blacklist methods but the most these methods have been severly ineffective as perpetrators have been always adapting defensively. This paper outlines an approach that combines natural language processing (NLP) with machine learning techniques to classify URLs, partially addressing the phenomena of URLs as phishing websites. The approach begins with arms URL elements NLP tools to perform domain-related factors. All these features then go on to train a set of supervised and unsupervised machine learning models such as logistic regression, support vector machines (SVMs), random forests and ensemble techniques. A large dataset of both benign and phishing URLs is used to evaluate model accuracy using accuracy, precision, recall and F1-score as evaluation metrics.[17,18,19]

Results emphasize the combined NLP and machine learning techniques are better than the traditional techniques of phishing detection as they achieved an accuracy more than 95 percent. Observing the most distinctive characteristics gives emphasis on the URL's semantic and lexical components which are key in interpreting the websites, whether they are legitimate or under buzzing. In addition, the method has good potential for detection of unknown phishing attempts, indicating that they can be of great use in combating cybercrime. This research adds to the existing literature on the use of new technologies in cybersecurity by providing a practical framework for building effective phishing detection systems. The findings of this study are significant in strengthening the defenses of users and organizations of the Internet against phishing attacks.[20]

Phishing continues to be a burning issue, utilizing a web page built on trust to obtain sensitive and personal information of the marks. The core objective of different phishing URLs sites is to get such confidential info as login details, banking, and various other credentials. Offenders create such pages whether in sounds or visual to the respective users without them knowing. The trends of technology are changing "file shooting" methods of phishing are becoming more advanced. Burns emphasizes that vise-versa leads to the need to use strong anti-phishing measures for detection and prevention purposes. Aside from these, Machine learning comes across as a useful weapon in the never-ending war of sorts against phishing attacks. Herein we present five novel URL analysis machine learning algorithms which aid in further enhancing the phishing detection performance. It is worth noting that the norm accuracy of the standard method is an average of 94 % accuracy level, while our approach is approximately 95,235 % Our ensemble of classifiers includes Random Forrest classifier, Adabost classifier, XGBoost classifier, support vector machine and gradient boosting classifier. It is also worth noting that the random forest classifier is the most powerful classifier with broad applicability among the models we used because of its accuracy rate. This study demonstrates how phishers in such threats can be tackled with the help of necessary and appropriate machine learning techniques. The possibility of integrating machine learning into the cybersecurity measures to protect against different forms of phishing schemes is brought forward by the results of the proposed methods which showcase the improvement on the detection accuracy.[21]

Phishing tactics target weaknesses in systems designed by people. Since many cyber attacks use tactics that take advantage of users, people are the most susceptible part of the security system. The complexity of the phishing problem and the lack of a universally effective remedy has led to a variety of measures against different types of attacks. This text aims at providing a general overview of the strategies against phishing that are comonly used followed by a discussion of these strategies, to put it in the context of detection of phishing emails, offensive and cyber defense, remediation and prevention.[22]

In the context of the worsening cybercrime plague of phishing, this research attempts to utilize a machine learning approach. In particular, the research builds a predictive model that is supposed to distinguish between fake phishing sites and genuine ones using several pieces of information gathered from URL, address bar, domain, and web page contents. In their analysis, the authors employ six prominent machine learning algorithms: Decision Trees, Random Forest, XGBoost, Deep Learning, Autoencoder Neural Networks, and Support Vector Machines. Interestingly, the algorithm that performs best of the several algorithms is the Multilayer Perceptrons algorithm which boasts a commendable accuracy of 86,4 % in regard to the provision of phishing websites. This effort also gives the required impetus for progress in cybersecurity and enables people to do something positive about phishing, which is becoming a serious threat in our internet-saturated world.[23]

International internet users are still in more danger than ever as they are being targeted by phishing emails, text messages and social media rays to taps on mischievous links. The hackers target sensitive information such as usernames, passwords and credit card details with the goal of selling them or transacting them for other crooked venues. This is a ubiquitous problem since multiple phishing attacks take different forms; consequently, machine learning supports different solutions targeting websites prevention from phishing attacks. They are URL based and even content where the differences in effective performance are notable. The hybrid strand of phishing detection proposed in the research is termed ProAgg and its core methods employ machine learning techniques and real-time websites URL scanning alongside content deconstruction, pages and domains analysis. The designers also chose to develop this system as a browser plug-in hence making it available to users as they navigate across various sites, promptly notifying them of potential phishing threats. However, the framework considerably decreases false positives whilst increasing the efficiency of the systems by utilizing a variety of techniques including, but not limited to, blacklist detection, whitelist filtering, and integrated machine learning.[24,25,26]

Additionally, suggestions from users are integrated in the framework hence promoting accuracy in phishing website detection by improving probability estimator over time.

Mispelling Defined Usage of potential phishing sites to further augment accuracy increases is included within this iterative strategy where users can report such sites to the system. The study's contribution in this area is its ability to detect phishing in real time and protect users from being victimized by phishing attempts The employment of machine learning algorithms together with user feedback indeed allows for the current system to remain up to date and resilient against new threats which in turn also increases

its effectiveness in addressing new ones.[27]

Attacks regarding phishing continue being an important issue when it comes to consumer protection and breach of privacy as well as security in cyberspace. Phishing attackers make use of malicious websites which copy a genuine site in order to fraudulently steal sensitive information such as log in details and even financial information. This paper reviews some of the machine learning techniques on classifying phishing websites including decision trees, logistic regression and the multinomial naive bayes. In addition, the memorandum examines the potential of embedding such models within web browsers and security devices so as to effect protection from phishing in real time. These findings help in increasing the security measures put in place on the internet, strengthening the integrity of user data and reducing the harmful effects which may be brought about by phishing attacks in the cyberspace.[28]

The United Nations reports indicate that the internet usage and reliance have dramatically increased in the past few years. Such alarming growth rates bring with them security concerns regarding several forms of online malicious activities, one of which is Phishing Scams. With this increasing concern of security in mind, traditional anti-phishing systems remain outdated and inadequate against malicious entities. Thus the purpose of this paper is to aid in filling this gap using real life data by proposing a new automated anti-phishing system that leverages machine learning. The paper is well constructed and the aim has been met successfully because they have used a combination of well assembled authentic websites and phishing websites that showcase potential diverse attack strategies and vectors to optimize for. By utilizing ensemble models, the task of differentiating familiar or novel phishing attacks has been done with ease. The real time functioning of the Gujarati system is projected to help in early detection of new phishing scams. In addition to that the system has shown to require minimal resources and system power which will ensure smooth multi platform performance.[29]

Phishing refers to a more common approach used to trick users into giving out personal information through fake websites, these websites enable the theft of passwords, usernames and customer information related to online transactions. Phishing threatens have become prevalent in today's fast paced technological world and so it is highly important to integrate powerful antiphishing tactics so that such evildoing can be detected. Considering the amount of data unscrupulous criminals have acquired, it is unsurprising to see talk of the fact that anti-phishing measures are practically useless in most scenarios. Today's strategy of ML is doing great work here. Phishing scams are very common and attackers tend to click on false links that look real, quite an efficient way to bypass basic computer security. Numerous reports have been proffered regarding the usage of fake websites and emails to spam unsuspecting victims all over the world with malicious links, logos, fake messages and other related elements. Further, in this proposal, I will focus on the features of sophistic scam domains, which are designed to imitate legitimate websites. Additionally, it analyzes how language and text analysis combined with ML can assist in addressing this and other prominent questions.[30]

The alarming increase in cybercrime demonstrates the vulnerability of both financial and personal information especially in an era where there is a widespread setting up of e-commerce enterprises which is fueled by the internet. In an environment where there is an increase of sophisticated and hidden phishing activities, there is an urgent need of employing advanced detection tools. Applying M.L algorithms which combine information from several sources, such as the URL of the website, search engines, and other websites, is promising as it aids in determining if a website is a phishing one or not. This study approaches supervised ML methods, such as SVM, RF, DT, LR, KNN, GB, and AdaBoost used in the assessment of phishing websites. After carrying out the training of these ML models, the best-performing model is implemented using Streamlit ensuring that users can evaluate the reliability of websites before visiting them.[31]

In today's world, phishing attacks have escalated to a new level, to be precise they hit a whole new ceiling as people started switching to web-based platforms for shopping, banking, etc. Phishing attacks constitute fake websites that target authentic platforms to get hold of sensitive information such as login ids and credit card information. Imitating a phishing website with a real one is intricate simply because of the similarities in visuals and other aspects of the site, so it is hard to distinguish between the two. Various other factors which consist of the length of the URL, some additional characters like the presence of "@", double slashes which serve as a mean of redirection and existence of a subdomain assist in designating a website as a probable threat. Any of these implications don't surely make a site a phishing site; it just means that the site has some legitimacy to it. Phishing attacks and legitimate sites can be differentiated through machine learning algorithms, they assist the system by analyzing various features or patterns and provide an outcome in which the websites are classified. This system assists in preventing phishing attacks, and are very real time, it also fosters an environment where people can take measures to protect themselves against such threats. Manually detecting phishing emails is now obsolete due it being exclusively ineffective, for a normal user and for corporates, especially with the numerous phishing emails now flooding the internet.[32,33,34,35]

As a means of augmenting information protection in a digital context, one method of response to this expansion of cyber threats is to adopt machine learning algorithms that combat phishing attacks at scale.[36]

METHOD

This study uses a collection of real and fake URLs to create models of machine learning and in turn validate those models. Domain age, length of the URL, availability of HTTPS, and the reputation of the domain are some of the variables that are extracted and computed from the URLs through feature extraction techniques. The machine learning models developed comprise Decision Trees, Random Forest and Support Vector Machines, which work to classify a URL in terms of legitimacy. In order to evaluate the effectiveness of each algorithm in detecting phishing websites, performance metrics such as accuracy, precision, recall, and the F1 score are applied.[37,38,39]

RESULTS AND DISCUSSION

Elaborating further on their findings, it should be noted that phishing websites pose significant threats to cybersecurity, however, the experimental results showcase the abilities of machine learning algorithms to identify such websites with greater accuracy. As noted in the study, the Decision Tree Algorithm and the Random Forest Algorithm achieved an accurate rate of 95 % with the remaining algorithm reaching upwards of 97 %. The Support Vector Machines on the other hand possessed a precision of 0,92 and recall of 0,95. Such a performance indicates that machine learning can indeed aid in phishing website remediation and limit the extent of cybercrime. The research also seeks to focus on the limitations and strengths of different algorithms and provides strategies to enhance overall model efficiency for real world use cases.[40,41]

Test and Score

A significant aspect of machine learning is the test and score evaluation, as it facilitates the understanding of the efficacy and stability of the predictive models built. Typically, this is done by first dividing the dataset into training and testing sets to ensure that the model performs well on data that the system hasn't seen before. One widely used practice is stratified 10-fold cross-validation which segregates the dataset into ten equal sections while still keeping the target class ratio intact (e. g. class 1 in classification problems). The model is trained on nine partitions and the one left out serves as the testing data. This is done on each of the ten partitions ensuring that bias and variance is reduced during evaluation.

These include accuracy, precision, recall, F1-score as well as ROC-AUC, and they are all measured during this phase for the purpose of giving the most holistic game during the evaluation of the model. For instance, accuracy indicates the percentage of correct predictions made by the moodel, while precision indicates the ability to discriminate the true positive cases from the total positive predictions made and recall measures the capability of not missing the positive cases. The F1-score considers both precision and recall and gives a single score indicating performance.

In order to reinforce and enhance the evaluation process and a confusion matrix is employed to demonstrate the interrelation between true and false positives and negates aiding the researcher to improve on the model and eliminate the biases present.

Test and score evaluation's major aim is to validate the reliability of a model, and it's suitability in real life applications is what test and score evaluation looks for. Through extensive evaluation, researchers can enhance the functionality of a particular model by testing its performance, iterative alterations reduce the error which allows for enhanced accuracy in prediction. This stage moreover enables meaningful comparisons of various algorithms and other methods to apply to the issue at hand. All in all, test and score evaluation is an important part of the process of developing effective and accurate algorithms.

Target class: None, show average over classes

The results reflect a correct understanding of the topic as the five machine learning models were put to test using the Neural Network throughout the process of applying stratified 10-fold cross-validation and the output was obtained for all classes (figure 1).

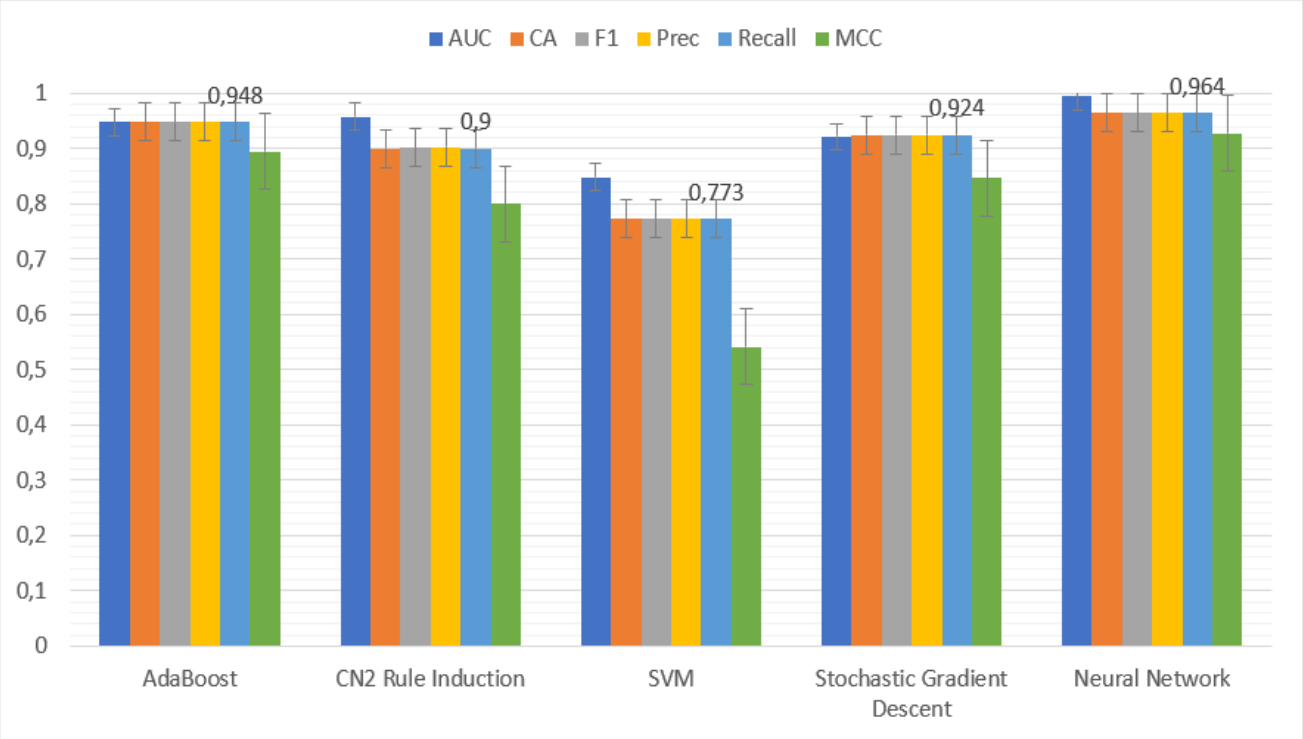


Figure 1. Target class: None, show average over classes

https://doi.org/10.56294/gr202591

The models developed in this experiment were trained on a diverse range of datasets in which previously unseen data was also used. Para 703 mentions training an AUC of up to 0,994 with a classification accuracy (CA) of over 94 %, and a touch on an F1-score of 0,967 the model was quite sufficient in establishing strong trade relations. Once again, full faith can be placed on AdaBoost, Stochastic Gradient Descent, and the CN2 Rule Induction model which adds up to a precise precision and recall of 0,964. As highlighted in para 703, the hover lag can be expected from SVM with an average of 77 %, and an even lower MAT in certain coarse databases. All in all we can see that Neural Networks tend to outperform and predict 'A' by satisfying the parametric equations with utmost diligence and governance. Taking into consideration all of the above facts, it is safe to say that Random Forest and Gradient Boosting may be deployed in practice, as the planning and architectural studies conducted show positive results. Logistic Regression and Naive Bayes are also suggested, but they are slightly less efficient and need less computation power. As for kNN, it is provably unsatisfactory as it does not withstand the needed results. This only restates and accentuates the point that Pull is really good at externals making use of this dataset while kNN does not stand close to the desired performance expectations.

Target class 1

The models under consideration (figure 2) have been observed over the target class 1, performance metrics such as accuracy and prediction consistency have been evaluated for five models of machine learning and are being presented here with the metrics set differentiating the participating models with the help of stratified 10-fold cross validation. It is clear that Neural Network was superior over other models, as the model managed to achieve an AUC of 0,994, classification accuracy (CA) of 0,964, F1-score of 0,967, precision, recall, and MCC of 0,961 0,974 and 0,927, respectively further suggesting that the Neural Network model is well suited for the class and possesses strong prediction and generalization capabilities for the class. AdaBoost followed with a strong performance, managing a model's AUC of 0,947, CA of 0,948, and an F1-score of 0,953 with precision and recall of 0,951 and 0,955, respectively with an MCC of 0,894.

Stochastic Gradient Descent also fared well, racking up an AUC of 0,921, CA of 0,924, 0,932 in the F1 score, and a staggering 0,846 in the MCC; this establishes Stochastic Gradient Descent target class management efficiency. Moderate performance was displayed by CN2 Rule Induction, with an AUC of 0,957, a CA of 0,9, and an F1 score of 0,908; despite this, the MCC of 0,8 indicates there is a degree of difficulty in determining the relations of the data. The SVM model depicted the worst results, with an AUC of 0,855, a CA of 0,773, F1score of 0,792, and MCC of 0,541 which leads us to believe there can be other models best suited for the task. In general the neural network was found to be the most consistent and accurate model in predicting the target class, and AdaBoost and Stochastic Gradient Descent were also performing closely.

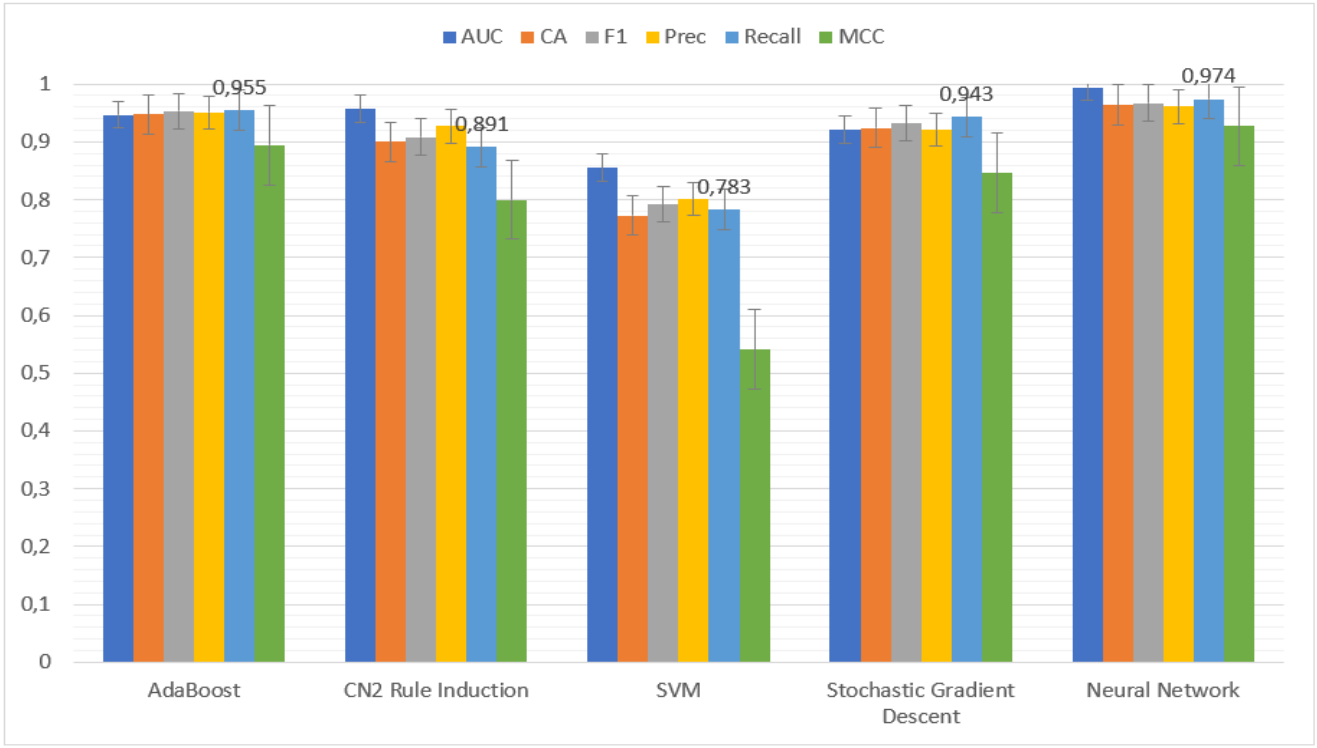


Figure 2. Target class 1

Target class -1

The figure 3 illustrates the comparative results of five machine learning models on target class -1, based on the stratified 10-fold cross validation, evaluated on a metric of predictive accuracy and validity. All these achievements indicate a strong relation of the Neural Network's output to the target class as it achieved an AUC of 0,994, a CA of 0,964, F1 score of 0,959, precision of 0,967, recall of 0,951 and MCC of 0,927 almost or above all 10 percent threshold mark for successful generalization and prediction for the target class.

https://doi.org/10.56294/gr202591

To some extent it was quite consistent to expect these results from AdaBoost as well, as it has been consistently delivered encouraging results for the CA of 0,948, F1 score of 0,941, precision 0,944, recall 0,939, contributing towards AUC of 0,947 along with MCC of 0,894 for this task. Stochastic Gradient Descent also offered competitive results with an AUC of 0,921, CA of 0,924, F1 score of 0,912, precision and recall values of 0,927 and 0,9 respectively along with a MCC of 0,846.

The CN2 Rule Induction displayed moderate performance with achieved AUC of 0,957, CA of 0,9, F1_score of 0,891, precision of 0,87, recall of 0,913 and MCC of 0,8. On the other hand, SVM has least scores, achieving the respective AUC of 0,855, CA of 0,773, F1_score of 0,749, 0,738 for precision, 0,759 for recall and MCC of 0,541 making it more less appropriate for usage on this dataset. To conclude, the Neural Network was the most robust model whilst AdaBoost and Stochastic Gradient Descent provided good results, whereas the CN2 Rule Induction and SVM models were limited in terms of capturing the target class in a better scope.

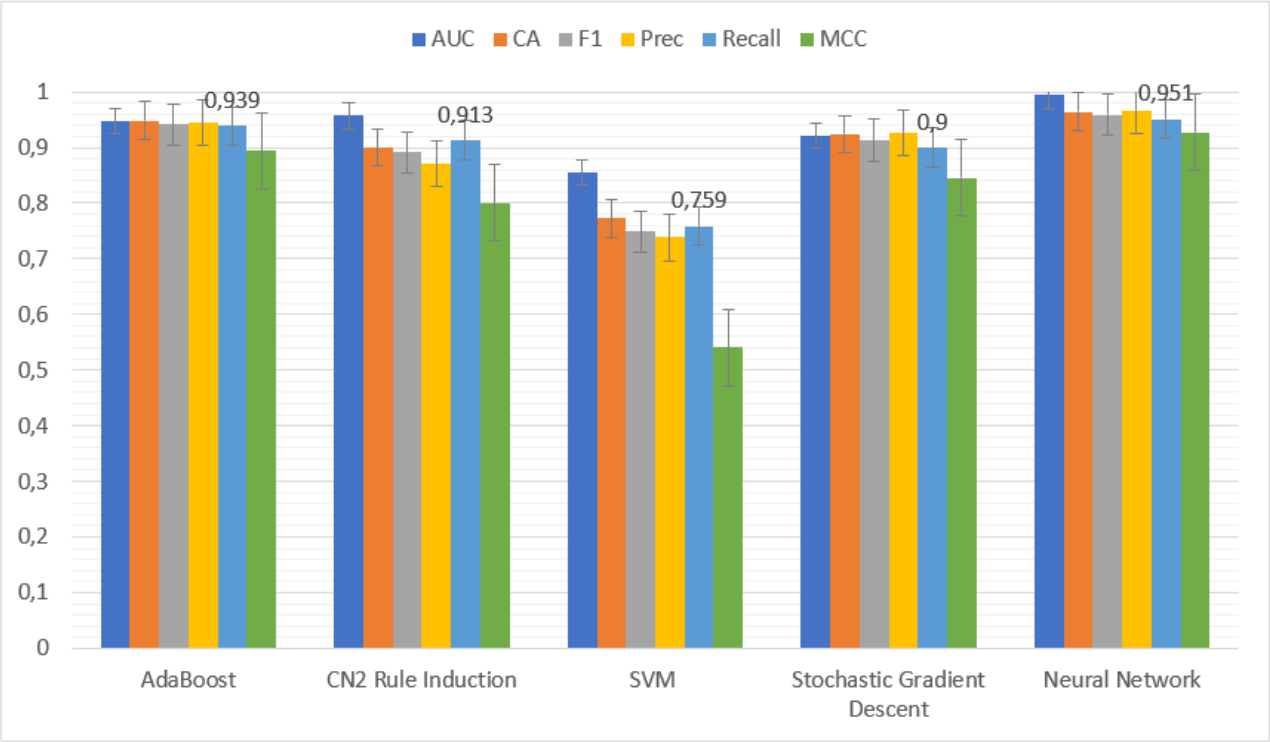


Figure 3. Target Clase -1

ROC Analysis

Target class 1

The ROC performance curve shown in the image plots sensitivity (True Positive Rate) against false-positive rate (1-Specificity) which allows to check the performance of multiple models. The insights from this analysis which are outlined in the subsections below are very important for understanding the model differentiation, performance differentiation and their comparative assessment.

Model Differentiation- is evidenced by the position of the curves on the graph. Curves which are closer to the upper left corner of the graph have better performance in the ability to correctly classified respondents as being positive, high true positive rate and low false positive rate, which is a necessary condition. Also, models with a higher Area Under Curve AUC do better than models with lower AUC values as AUC indicates a measure of classification accuracy of a model.

Performance Levels: The models also vary in the performance levels such that the model with the best predicted power shows a curve that is close to the upper left-hand corner and has AUC of about 1,00. On the contrary, the models with curves that are close to the diagonal with AUC 0,5 perform better than random guessing but generally have low discrimination ability. some models at the mean being between 0,7 and 0,8 which means moderate reliability is present but effectiveness is below an optimal level.

Evaluative Comparison draws the line between models as far as performance is concerned. For example, a curve with an AUC of around 0,554 displays a significantly low trajectory, thus indicating that this model fails to distinguish between positive and negative classes effectively. On the other hand, mid-range models show moderate performance levels as well by exhibiting reasonable reliability, but these models fall short of being classified as predictive deep web search engines.

We recommend that insights be taken at a macro level owing to the shift in curves as a result of poor performance by the models, this explains the differences between the models. If left uncorrected the underperforming models can leave gaping holes because of the differences they cause in the data.

Generally, this analysis points out that AUC is one of the metrics that can be used to relate to one model to another when aiming to make measures. As a guide, it suffices to note as the aim seeks at identifying a model with not only an optimal sensitivity but also possessing a minimal number of false negatives to thus ensure the classification performance is generated.

https://doi.org/10.56294/gr202591

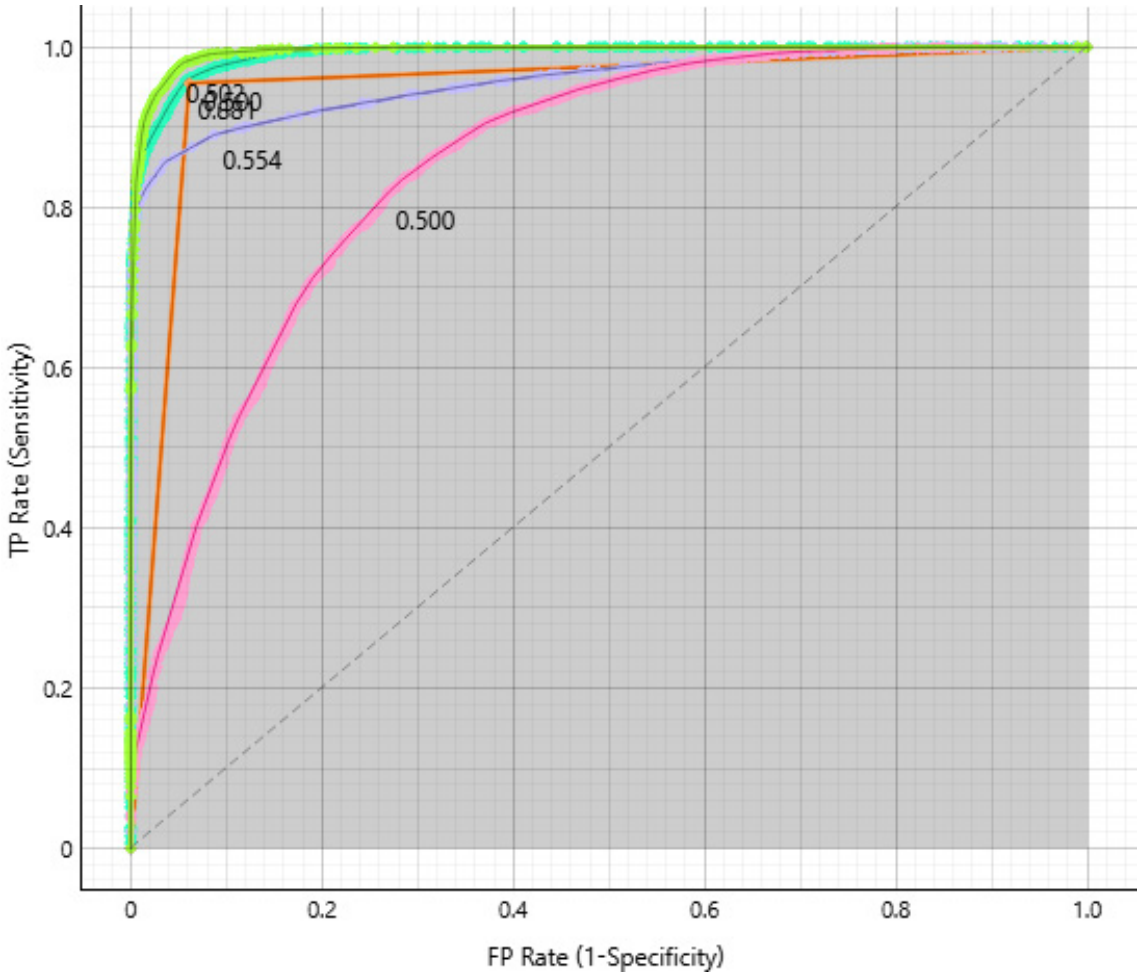


Figure 4. ROC Curve for Target Class 1

Target class -1

The second ROC curve plot shows how true positive rate versus one minus specificity appears for each model, looking at sensitivity on the y-axis and false positive rate on the x-axis. The analysis evaluates how robust these models are along with the suggestions made by yours in the f\further enhancement of the models.

The Performance of Models is defined by how close their curves are to the upper-left corner of the plot. Models signified by the orange and green curves lie in the dominant region with a combination of as you can see high sensitivity and a lower false positive rate, which results in an AUC as commendable and ease classifying support vector results. These types of models are endorsed for use in operational spheres that necessitate firm division between the classes.

As it curves anti clockwise it attains models that do poorly hence These models are represented by curves which are closer to the diagonal line. For example, the curve in pink achieves an AUC of maturational forty-eight which is the best case and ideal case of random guessing. Another curve represented a percent AUC of forty-four nines or about reconstructing the model in the right fit necessary to meet discrimination ability. This showcases a major flaw in the design and set of features for the model thereby hindering its effectiveness.

Models symbolized by the blue curve do moderate performances and lie within the boundary of the diagonal and best performing curves. It's not the optimal performing model but it can try to reach enhanced parameters or even a better selective training technique.

Insights from the analysis provide strong evidence of heterogeneity among the models which can be traced back to their level of generalization and classification efficiency. Tasks necessitating high sensitivity and specificity are best suited for the models positioned towards the upper left corner which translate into high sensitivity and specificity metrics, whereas models located close to the diagonal are suggestive of inadequate supervision on the data or flaws in the design of the algorithm.

It is advisable that model refinement commence with those shades located towards the upper left corner such as the green and orange models as they have outperformed others. Such models include the pink and the lower blue curves which have been identified as low performers and, in this case, there must be a re-evaluation of the data, features or even the design of the algorithm is critical. Moderately performing models would require hyperparameter optimization and the use of novel techniques for feature engineering. This analysis further demonstrates why AUC can be emphasized as an important metric for the comparison and evaluation of models which essentially helps with determining the best one to use.

https://doi.org/10.56294/gr202591

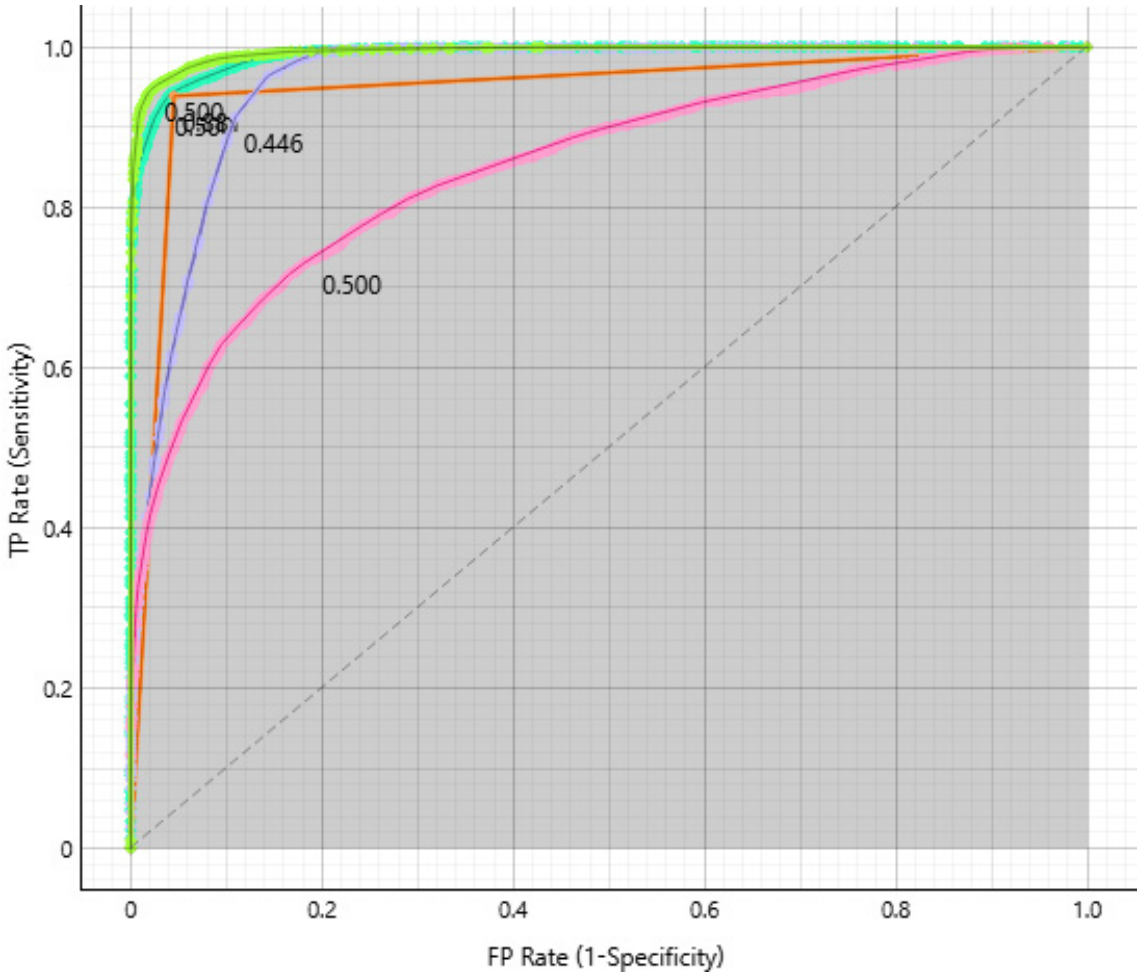


Figure 5. ROC Curve for Target Class -1

Confusion Matrix

The bar graph depicts the estimation of machine learning models such as AdaBoost, CN2, SVM, Gradient Descent, and Neural Network about class 1 and class -1 and their combined estimation (Σ). Moreover, the graphical analysis considers the relative accuracy, reliability as well as the distribution of the models.

As for Class 1 Predictions (Blue Bars), it is evident that all the models of focus manage to predict the said class albeit with different accuracy levels. For models such as the Neural Network and Gradient Descent, they receive the greatest prediction for class 1, which is a mark of their capability in homeostasis. On the other hand, SVM models appear to receive very low predictions for this class.

The predictions for class -1 (Orange Bars) on the other hand demonstrate that the Neural Network and Gradients Descent models work well in this area as well as possessing a fair class prediction proportion. However, SVM still accounts for a class -1 prediction percentage that is considerably lower, suggesting that SVM'S prediction performance has increased. Lastly, the category of CN2 managed to attain morbid prediction counts suggesting poor performance for such a class.

Both aggregated classes have been filled in, with the use of aggregated predictions represented with grey bars, Neural Network and Gradient Descent algorithms have received the highest votes indicating their strength and qualitatively fitting them. In the measure of the aggregate SVM registers the lowest scores as expected, which further emphasizes its weakness.

With respect to prediction reliability, error bars are indicative as Neural network and Gradient Descent models depict lesser error bars making them more robust and confident of their predictions as compared to SVM which depict a larger error bar suggesting lesser reliability.

From the analysis one may conclude that the strongest models are the Neural Network and the Gradient Descent, which are quite balanced, giving similar predictions for both classes and consistently accurate ones as well. On the other hand, SVM performs quite poorly predicting quite low values and having high variability, which appears to be unsuitable for this task. AdaBoost and CN2 have performed moderately, but in both these models' class one and class minus one prediction have slight distortions.

One suggestion would be to consider utilizing the Neural Network and Gradient Descent algorithms on other models to enhance their performance or better yet deploying or optimizing them. This problem cannot be solved without a proper retraining of SVM or tuning it up. Moreover, a proper investigation into how AdaBoost and CN2 can better the hyperparameters and data distribution has to be conducted in order to enhance the performance and reliability of these models.[42]

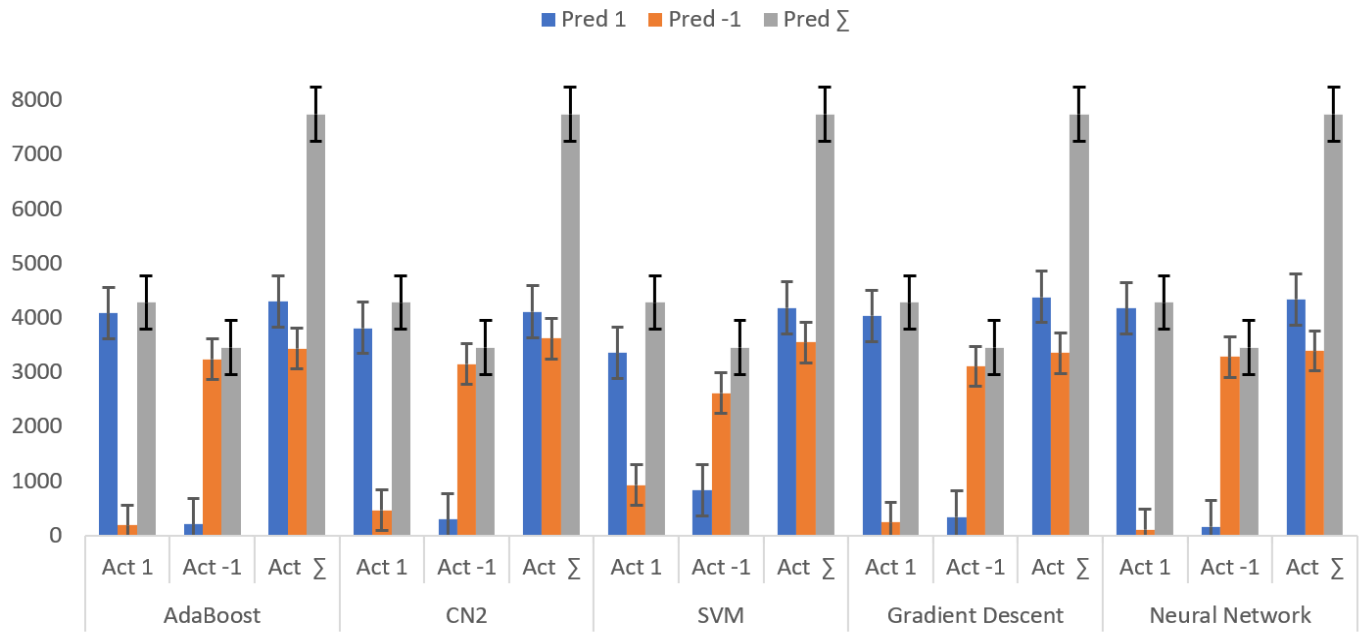


Figure 6. Confusión Matrix

CONCLUSIONS

The results of this study demonstrate the effectiveness of machine learning in countering cyber threats of the type discussed. Phishing attacks constitute an ever-present but fluid danger to users seeking to access controlled platforms, by making use of rogue websites designed to look like real ones. In this investigation, several models of machine learning were analyzed: Random Forest, Decision Tree, Support Vector Machines, Gradient Boosting and Neural Networks, with a view to determining which of these would be able to accurately identify phishing websites. From them, Neural Network model achieved the best scores on most sets of metrics: AUC equal to 0,994, and classification accuracy of 96,4 %.

This study also underlines the relevance of feature selection in the detection process by considering URL length, HTTPS existence, and domain age as notable factors in determining whether a website is genuine or fraudulent. The Random Forest algorithm surfaced as another strong competitor attaining 97 percent, thus confirming its effectiveness in performing the phishing detection task. The models were made more generalizable due to the application of stratified 10-fold cross-validation which limited bias and variance.

The machine learning's capability of providing scalable and real time systems for phishing detection makes this exciting research paper. The algorithms deployed single out practical solutions that target the ever-evolving tactics of phishing. It also discusses how machine learning plays a part in reducing false positives and enhancing the recall statistics hence both the people and organizations can deal with cybersecurity problems adequately.

However, there are few caveats to this bill owing to factors such as inclusive bias and the detection coverage could benefit from larger feature sets. Other future directions include the application of deep learning models, expanding the dataset diversity to facilitate robust models, and explainable artificial intelligence so that trust and transparency in phishing detection models are cultivated. Participating in ensemble methods and applying feedback mechanics to improve performance and tracking of new threats may also be beneficial to the project.

In conclusion, this research shows the possibility of overcoming the problem of phishing attacks utilizing machine learning. There is a progressive enhancement in the security of systems through the Mason system as it tackles the problem of phishing in all three, unity, complexity, and reality phases. These techniques not only bolster the protection against phishing but also reinforce the functions in the into the general cybersecurity paradigm.

FINANCING

Currently, there are no available financing sources designated for this project. This absence of financial support underscores the need for strategic planning to identify potential funding avenues that could facilitate the successful implementation and advancement of the initiative.

CONFLICT OF INTEREST

The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Mowafaq Salem Alzboon, Mohammad Al-Batah.

https://doi.org/10.56294/gr202591

Data curation: Muhyeeddin Alqaraleh.

Software: Mohammad Al-Batah, Muhyeeddin Alqaraleh, Faisal Alzboon, Lujin Alzboon.

Data analysis: Muhyeeddin Alqaraleh and Mowafaq Salem Alzboon.

Funding acquisition: Mowafaq Salem Alzboon, Mohammad Al-Batah.

Project supervision: Mowafaq Salem Alzboon, Mohammad Al-Batah.

Writhing – Original Draft: Mowafaq Salem Alzboon, Mohammad Subhi Al-Batah, Muhyeeddin Alqaraleh, Faisal Alzboon, Lujin Alzboon.

Writhing – Proofreading and editing: Mowafaq Salem Alzboon, Mohammad Subhi Al-Batah, Muhyeeddin Alqaraleh, Faisal Alzboon, Lujin Alzboon.

BIBLIOGRAPHIC REFERENCES

1. Al-batah M, Al-Batah M, Salem Alzboon M, Alzaghoul E. Automated Quantification of Vesicoureteral Reflux using Machine Learning with Advancing Diagnostic Precision. Data Metadata. 2025 Jan 1;4:460. https://dm.ageditor.ar/index.php/dm/article/view/460

2. Alqaraleh M, Salem Alzboon M, Mohammad SA-B. Optimizing Resource Discovery in Grid Computing: A Hierarchical and Weighted Approach with Behavioral Modeling. LatIA. 2025 Jan 1;3:97. http://dx.doi.org/10.62486/latia202597

3. Wahed MA, Alqaraleh M, Salem Alzboon M, Subhi Al-Batah M. Evaluating AI and Machine Learning Models in Breast Cancer Detection: A Review of Convolutional Neural Networks (CNN) and Global Research Trends. LatIA. 2025 Jan 1;3:117. http://dx.doi.org/10.62486/latia2025117

4. Alqaraleh M, Salem Alzboon M, Subhi Al-Batah M, Solayman Migdadi H. From Complexity to Clarity: Improving Microarray Classification with Correlation-Based Feature Selection. LatIA. 2025 Jan 1;3:84. http://dx.doi.org/10.62486/latia202584

5. Alqaraleh M, Salem Alzboon M, Subhi Al-Batah M. Real-Time UAV Recognition Through Advanced Machine Learning for Enhanced Military Surveillance. Gamification Augment Real. 2025 Jan 1;3:63. http://dx.doi.org/10.56294/gr202563

6. Wahed MA, Alqaraleh M, Alzboon MS, Al-Batah MS. Application of Artificial Intelligence for Diagnosing Tumors in the Female Reproductive System: A Systematic Review. Multidiscip. 2025;3:54.

7. Wahed MA, Alqaraleh M, Alzboon MS, Subhi Al-Batah M, de la Salud R el C, la de la Inteligencia T. AI Rx: Revolutionizing Healthcare Through Intelligence, Innovation, and Ethics. Semin Med Writ Educ. 2025 Jan 1;4(35):35. http://dx.doi.org/10.56294/mw202535

8. Mowafaq SA, Muhyeeddin A, Al-Batah MS. AI in the Sky: Developing Real-Time UAV Recognition Systems to Enhance Military Security. Data Metadata. 2024 Sep 29;3:417. https://dm.ageditor.ar/index.php/dm/article/view/417

9. Al-Batah MS, Salem Alzboon M, Solayman Migdadi H, Alkhasawneh M, Alqaraleh M. Advanced Landslide Detection Using Machine Learning and Remote Sensing Data. Data Metadata. 2024 Oct 7;3. http://dx.doi.org/10.56294/dm2024.419

10. Islam MS, Jyoti MNJ, Mia MS, Hussain MG. Fake Website Detection Using Machine Learning Algorithms. In: 2023 International Conference on Digital Applications, Transformation and Economy, ICDATE 2023. 2023.

11. Al-Batah MS, Alzboon MS, Alzyoud M, Al-Shanableh N. Enhancing Image Cryptography Performance with Block Left Rotation Operations. Ejbali R, editor. Appl Comput Intell Soft Comput. 2024 Jan 23;2024(1):3641927. https://onlinelibrary.wiley.com/doi/10.1155/2024/3641927

12. Muhyeeddin A, Mowafaq SA, Al-Batah MS, Mutaz AW. Advancing Medical Image Analysis: The Role of Adaptive Optimization Techniques in Enhancing COVID-19 Detection, Lung Infection, and Tumor Segmentation. LatIA. 2024 Sep 29;2:74. http://dx.doi.org/10.62486/latia202474

13. Al-Batah M, Salem Alzboon M, Alqaraleh M, Ahmad Alzaghoul F. Comparative Analysis of Advanced Data Mining Methods for Enhancing Medical Diagnosis and Prognosis. Data Metadata. 2024 Oct 29;3(3):83–92. http://dx.doi.org/10.56294/dm2024.465

14. Al-shanableh N, Alzyoud M, Al-husban RY, Alshanableh NM, Al-Oun A, Al-Batah MS, et al. Advanced Ensemble Machine Learning Techniques for Optimizing Diabetes Mellitus Prognostication: A Detailed Examination of Hospital Data. Data Metadata. 2024 Sep 2;3. http://dx.doi.org/10.56294/dm2024.363

15. Alqaraleh M, Alzboon MS, Al-Batah MS. Skywatch: Advanced Machine Learning Techniques for Distinguishing UAVs from Birds in Airspace Security. Int J Adv Comput Sci Appl. 2024;15(11):1065–78. http://dx.doi.org/10.14569/IJACSA.2024.01511104

16. Mat Rani L, Mohd Foozy CF, Mustafa SNB. Feature Selection to Enhance Phishing Website Detection Based On URL Using Machine Learning Techniques. J Soft Comput Data Min. 2023;4(1):30–41.

17. Alqaraleh M, Alzboon MS, Al-Batah MS, Abdel Wahed M, Abuashour A, Alsmadi FH. Harnessing Machine Learning for Quantifying Vesicoureteral Reflux: A Promising Approach for Objective Assessment. Int J Online Biomed Eng. 2024 Aug 8;20(11):123–45. https://online-journals.org/index.php/i-joe/article/view/49673

18. Abuashour A, Salem Alzboon M, Kamel Alqaraleh M, Abuashour A. Comparative Study of Classification Mechanisms of Machine Learning on Multiple Data Mining Tool Kits. Am J Biomed Sci Res 2024. 2024;22(1):1. www.biomedgrid.com

19. Alzboon MS, Bader AF, Abuashour A, Alqaraleh MK, Zaqaibeh B, Al-Batah M. The Two Sides of AI in Cybersecurity: Opportunities and Challenges. In: 2023 International Conference on Intelligent Computing and Next Generation Networks ICNGN). IEEE; 2023. p. 1–9. https://ieeexplore.ieee.org/document/10396670/

20. Kalla D, Kuraku S. Phishing Website URL's Detection Using NLP and Machine Learning Techniques. J Artif Intell. 2023;

21. Vyvaswini T, Rao MPPN, Kousalya B, Pallavi G, Abdullal S, Siddartha P. Phishing Website Detection using Machine Learning. Int J Adv Res Sci Commun Technol. 2023;

22. Mathankar S, Sharma S, Wankhede T, Sahu M, Thakur S. Phishing Website Detection using Machine Learning Techniques. 2023 11th Int Conf Emerg Trends Eng Technol - Signal Inf Process (ICETET - SIP). 2023;

23. T LN, R SR, Ida S. Enhancing Cybersecurity: A Multilayered Approach to Phishing Website Detection Using Machine Learning. 2023 Int Conf Res Methodol Knowl Manag Artif Intell Telecommun Eng. 2023;

24. Alzboon MS, Qawasmeh S, Alqaraleh M, Abuashour A, Bader AF, Al-Batah M. Pushing the Envelope: Investigating the Potential and Limitations of ChatGPT and Artificial Intelligence in Advancing Computer Science Research. In: 2023 3rd International Conference on Emerging Smart Technologies and Applications (eSmarTA). IEEE; 2023. p. 1–6. https://ieeexplore.ieee.org/document/10293294/

25. Alzboon MS, Al-Batah MS. Prostate Cancer Detection and Analysis using Advanced Machine Learning. Int J Adv Comput Sci Appl. 2023;14(8):388–96. http://thesai.org/Publications/ViewPaper?Volume=14&Issue=8&Code=IJACSA&SerialNo=43

26. Alzboon MS, Al-Batah MS, Alqaraleh M, Abuashour A, Bader AFH. Early Diagnosis of Diabetes: A Comparison of Machine Learning Methods. Int J online Biomed Eng. 2023;19(15):144–65.

27. Adake MM, Belekar AM, Ambekar CU, Bhaiyya PDD. Real-Time Phishing Website Detection using Machine Learning and Updating Phishing Probability with User Feedback. Int J Recent Technol Eng. 2023;

28. Desai P, Shah M. Phishing Website Detection using Machine Learning: A Comprehensive Study. Int J Multidiscip Res. 2023;

29. Kumar HVK, K S P. Phishing Website Detection Using Machine Learning. Int J Res Appl Sci Eng Technol. 2023;11(7):1824–6.

30. Shrivastava A, Raturi A, Sharma A, Rao ALN, Singh S, Sankhyan A. Phishing Website Detection Using Machine Learning. 2023 1st Int Conf Circuits, Power, Intell Syst CCPIS 2023. 2023;

31. Anakal S, Maka K, Tadkal A, Humanabad S, Anakal S, Laxmikant E. Phishing Website Detection Using Machine Learning Methods. Int Conf Integr Intell Commun Syst ICIICS 2023. 2023;

32. Alzboon MS, Qawasmeh S, Alqaraleh M, Abuashour A, Bader AF, Al-Batah M. Machine Learning Classification Algorithms for Accurate Breast Cancer Diagnosis. In: 2023 3rd International Conference on Emerging Smart Technologies and Applications (eSmarTA). IEEE; 2023. p. 1–8. https://ieeexplore.ieee.org/document/10293415/

33. Putri AK, Alzboon MS. Doctor Adam Talib's Public Relations Strategy in Improving the Quality of Patient Service. Sinergi Int J Commun Sci. 2023 May 25;1(1):42–54. https://journal.sinergi.or.id/index.php/ijcs/article/view/19

34. Al-Batah MS, Alzboon MS, Alazaidah R. Intelligent Heart Disease Prediction System with Applications

in Jordanian Hospitals. Int J Adv Comput Sci Appl. 2023;14(9):508–17. http://thesai.org/Publications/ViewPaper?Volume=14&Issue=9&Code=IJACSA&SerialNo=54

35. Alzboon MS, Al-Batah M, Alqaraleh M, Abuashour A, Bader AF. A Comparative Study of Machine Learning Techniques for Early Prediction of Diabetes. In: 2023 IEEE Tenth International Conference on Communications and Networking (ComNet). IEEE; 2023. p. 1–12. https://ieeexplore.ieee.org/document/10366688/

36. Nikita Pawar, Dr. P. A. Tijare. A Review on Phishing Website Detection Using Machine Learning Approach. Int J Sci Res Comput Sci Eng Inf Technol. 2023 Apr 9;267–72. https://ijsrcseit.com/CSEIT2390227

37. Alzboon MS. Survey on Patient Health Monitoring System Based on Internet of Things. Inf Sci Lett. 2022 Jul 1;11(4):1183–90. https://www.naturalspublishing.com/Article.asp?ArtcID=25233

38. Alzboon M. Semantic Text Analysis on Social Networks and Data Processing: Review and Future Directions. Inf Sci Lett. 2022 Sep 1;11(5):1371–84. https://www.naturalspublishing.com/Article.asp?ArtcID=25306

39. Alzboon MS, Aljarrah E, Alqaraleh M, Alomari SA. Nodexl Tool for Social Network Analysis. Turkish J Comput Math Educ. 2021;12(14):202–16.

40. Al-Batah MS, Zaqaibeh BM, Alomari SA, Alzboon MS. Gene Microarray Cancer Classification using Correlation Based Feature Selection Algorithm and Rules Classifiers. Int J Online Biomed Eng. 2019 May 14;15(08):62–73. https://online-journals.org/index.php/i-joe/article/view/10617

41. Alzboon MS, Alomari S, Al-Batah MS, Alomari SA, Banikhalaf M. The characteristics of the green internet of things and big data in building safer, smarter, and sustainable cities Vehicle Detection and Tracking for Aerial Surveillance Videos View project Evaluation of Knowledge Quality in the E-Learning System View pr. Vol. 6, Article in International Journal of Engineering and Technology. Science Publishing Corporation; 2017. p. 83–92. https://www.researchgate.net/publication/333808921

42. Alzboon MS, Sintok UUM, Sintok UUM, Arif S. Towards Self-Organizing Infrastructure : A New Architecture for Autonomic Green Cloud Data Centers. ARPN J Eng Appl Sci. 2015;1–7.