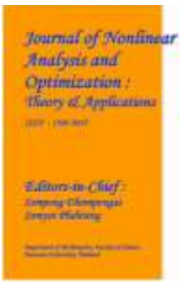
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**



DEEP LEARNING-BASED PHISHING EMAIL DETECTION USING MULTILEVEL FEATURE REPRESENTATIONS

#1.Mr. SHAIK HIMAM BASHA, #2.POTHULA DEEPTHI PRASANNA

#1.PG Scholar in the M.C.A QIS COLLEGE OF ENGINEERING & TECHNOLOGY

#2Assistant Professor Department of Master of Computer Applications

Vengamukkapalem(V), Ongole, Prakasam dist., Andhra Pradesh- 523272

Abstract

The phishing email is one of the significant threats in the world today and has caused tremendous financial losses. Although the methods of confrontation are continually being updated, the results of those methods are not very satisfactory at present. Moreover, phishing emails are growing at an alarming rate in recent years. Therefore, more effective phishing detection technology is needed to curb the threat of phishing emails. In this paper, we first analyzed the email structure. Then, based on an improved recurrent convolutional neural networks (RCNN) model with multilevel vectors and attention mechanism, we proposed a new phishing email detection approach that models emails at the email header, the email body, the character level, and the word level simultaneously. To evaluate the effectiveness of the approach, we use an unbalanced dataset that has realistic ratios of phishing and legitimate emails. The experimental results show that the overall accuracy is very high. Meanwhile, the false positive rate (FPR) is very low. High accuracy and low FPR ensure that the filter can identify phishing emails with high probability and filter out legitimate emails as little as possible.

Keywords — Phishing email detection, deep learning, RCNN, Attention mechanism, Multilevel feature representation, Email analysis, Cyber security.

Introduction:

Phishing attacks are a major threat to cyber security today. They trick users into giving away sensitive information like usernames, passwords, and bank details. These attacks usually come in the form of emails that look like they are from trusted sources.

Because phishing emails appear so real, they are hard to detect using normal spam filters or simple rule-based systems.As phishing tactics have become smarter and more complex, there is a strong need for better and more advanced detection methods.

Machine learning, especially deep learning, offers effective ways to fight phishing in real time.

Deep learning can learn patterns directly from data and adjust to new phishing tricks. One powerful approach is to use a combination of Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and attention mechanisms.

CNNs are great at identifying visual and structural features in emails, like layout and links . RNNs are useful for reading and understanding the sequence of words in the email . Attention mechanisms help the model focus on the most important parts of the email, like suspicious phrases or links . The Improved Recurrent Convolutional Neural Network (RCNN) model combines all three methods to give better results .

It uses CNNs to extract features, RNNs to understand the email flow, and attention to highlight the critical parts .

Phishing has changed a lot over time. Early phishing messages were easy to spot due to bad grammar and strange links . But now, phishing emails are well-written and use tricks like urgency or threats to fool people .

Phishing has also moved beyond emails and now appears in SMS, social media, and phone calls . This spread makes it harder to detect with old methods . Traditional systems rely on known rules and patterns, which means they often miss new types of phishing attacks .

They also produce too many false alarms. Deep learning systems can learn from data and adjust to new phishing styles without needing constant manual updates .These systems reduce errors and catch more threats.

Using advanced models like RCNN can help build strong, accurate phishing detection systems. These smart systems can protect users better and reduce the risks of cyber fraud.

With the growing number of phishing attempts, it is important to keep improving these models. Future tools will rely even more on deep learning to stay ahead of attackers.

Literature Survey:

1. Simon J. Bell & Peter Komisarczuk (2020)Title: Blacklist-Based Phishing Detection.Reference: https://www.researchgate.net/publication/338933845

Merits:

- Uses a blacklist of known phishing email addresses, domains, and URLs.
- Offers high accuracy and fast detection for previously reported phishing sources.
- Lightweight and efficient for resource-constrained environments.

Demerits:

- Ineffective against zero-day phishing attacks and obfuscated domains.
- Requires continuous updates and maintenance to remain effective.

2. Kayode S. Adewole & Abimbola G. Akintola (2019)Title: Rule-Based Email Filtering for Phishing Detection. Reference:https://www.researchgate.net/publication/334778644

Merits:

- Utilizes hybrid rule-induction algorithms (JRip and PART) for real-time filtering.
- Provides interpretable rule sets with reduced false alarm rates.
- Demonstrates high accuracy on publicly available datasets.

Demerits:

- Static rules lack adaptability to evolving phishing techniques.
- Can be bypassed by attackers using

3. Kumar & Rao (2022) Title: Phishing Email Detection Using Naïve Bayes Classification Reference: https://www.researchgate.net/publication/388324015

Merits:

- Simple and efficient probabilistic model.
- Achieves high accuracy ~97.82%) with minimal computational resources.
- Easy to implement for basic spam and phishing detection.

Demerits:

- Limited in handling modern and context-aware phishing techniques.
- Lacks deep semantic understanding of complex email content.

4. Prajakta Patil, Rashmi Rane & Madhuri Bhalekar (2017) Title: SVM-Based Email Phishing Detection Refer ence: https://www.researchgate.net/publication/320651414

Merits:

- Employs Support Vector Machine (SVM) with Map Reduce for scalability.
- Provides good accuracy with engineered feature sets.• Suitable for large-scale email datasets.

Demerits:

- Requires manual feature engineering and tuning.
- Higher computational load makes it less ideal for real-time applications.

5. Heena Kousar, Bibiana Jenifer, Afnan Khan et al. (2021)Title: Email Threat Detection Using RandomForest Reference: Based on "Phishing Email Detection Using Random Forest Technique"

Merits:

- Uses Random Forest for robust classification of phishing emails.
- Resistant to overfitting and performs well with diverse data features.
- Strong performance in practical phishing detection tasks.

Demerits:

- Requires significant memory and processing power.
- Less interpretable due to ensemble nature of the model.Implementation details and performance evaluations are not provided.
- Lacks comparative analysis with existing methods.ijitce.org.

Analysis:

An advanced phishing email detection system using RCNN and attention mechanisms analyzes emails at multiple levels (header, body, character, and word) to identify malicious intent, leveraging the power of RCNN for pattern recognition and attention mechanisms to focus on relevant parts of the email . This system examines emails at several levels, including the header, body, character sequences, and word-based details. RCNN detects hidden patterns in the text, helping identify harmful or Deceptive email structures. Attention mechanisms improve accuracy by targeting critical parts like dangerous links or abnormal sender information .This approach lowers error rates and strengthens the system’s ability to catch tricky or disguised phishing attempts. It learns from both basic and complex content, evolving as it processes more data. Over time, the system becomes more precise, making it a strong defense against phishing threats.

System Architecture:

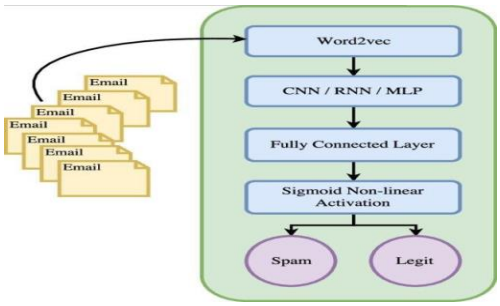


Fig 1: System Architecture

Modules:

User login: My details: It mean giving our information to receiver. Compose mail: The meaning of "compose mail" is simply the act of creating or drafting an email. It involves writing a message to a recipient using email tools. Typically, when you compose an email, you.

Checking phishing email details: "Checking phishing details" refers to the process of verifying whether certainInformation or communication is part of a phishing attempt. Phishing is a type of cybercrime where attackers deceive individuals into revealing sensitive information, such as passwords, credit card numbers, or personal data

View phishing details: Phishing is a type of cybercrime where attackers trick individuals into revealing sensitive information, such as passwords, credit card numbers, or personal data. Here are some key details to help you identify and understand phishing.

Feedback: Phishing attacks are a form of cybercrime where attackers impersonate trusted entities to deceive individuals into revealing sensitive information, such as passwords, credit card numbers, or personal data. Here's an overview.

Logout: "Logout" means ending your active session on a website, application, or system. By logging out, you disconnect your account from the device or platform you are using, ensuring that others cannot access your personal information or activities. It’s an important step for maintaining security, especially on shared or public devices. Would you like to know how to log out from a specific platform?

Algorithm:

R-CNN Algorithm: R-CNN Recurrent Convolutional Neural Network (R-CNN, Fast R-CNN, and Faster R-CNN) that we saw in the first article.

This will help lay the ground for our implementation part later when we

will predict the bounding boxes present in previously unseen images (new data).

R-CNN extracts a bunch of regions from the given image using selective search, and then checks if any of these boxes contains an object. We first extract these regions, and for each region, CNN is used to extract specific features. Finally, these features are then used to detect objects. Unfortunately, R-CNN becomes rather slow due to these multiple steps involved in the process.

Fast R-CNN, on the other hand, passes the entire image to ConvNet which generates regions of interest (instead of passing the extracted regions from the image). Also, instead of using three different models (as we saw in R-CNN), it uses a single model which extracts features from the regions, classifies them into different classes, and returns the bounding boxes.

All these steps are done simultaneou sly,thus making it execute faster as compared to R-CNN. Fast R-CNN is, however, not fast enough when applied on a large dataset as it also uses selective search for extracting the regions.

Implementation:

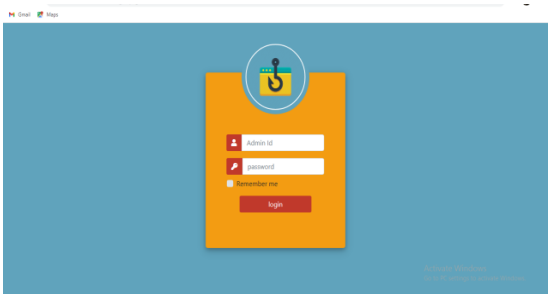


Fig 2: Home Page.

Description: The above screens the three modules i.e., Email, password, login.

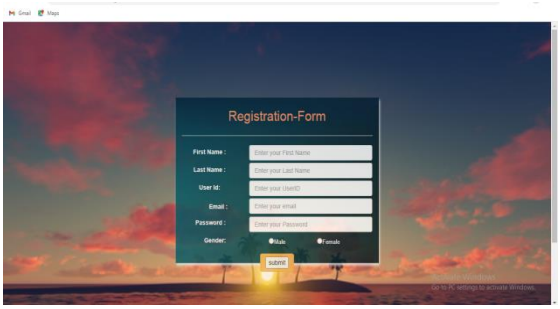


Fig 3: Sign Up.

Description: The above screen is used to register a person and all their details will be entered.

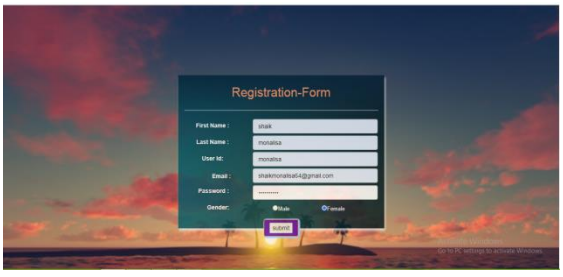


Fig 4: Registration form.

Description: The above screen a person enters the details to register.

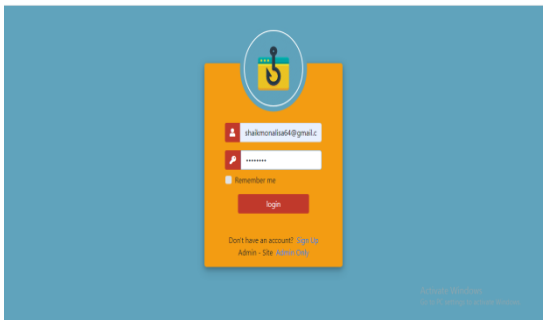


Fig 5: Login to phishing page.

Description: After the registration process login the phishing page.



Fig 6: Interface.

Description: This is a project interface.



Fig 7: Compose mail module.

Description: In this compose mail, the user can send mail to another person.

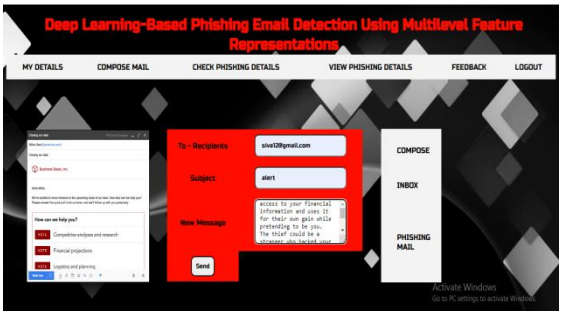


Fig 8: Compose mail module.

Description: In this compose mail sending to mail to another person.

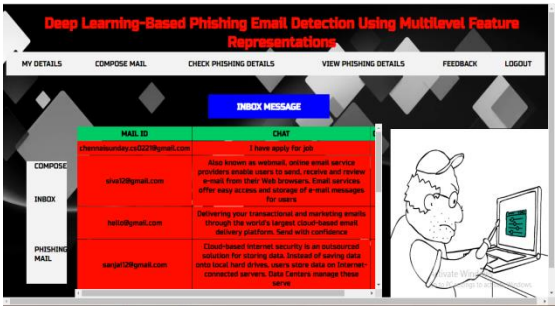


Fig 9: Compose mail module.

Description: In the compose mail has inbox mail that mean it is normal message.

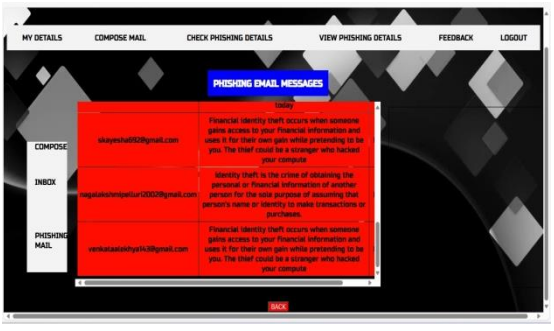


Fig 10: Compose mail module.

Description: In the compose mail it checks the it is phishing mail message it checks good message of phishing mail.

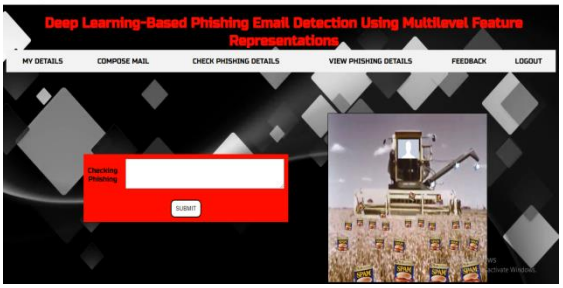


Fig 11: Compose mail module.

Description: phishing checking details means it checks the message it is normal or phishing attack.

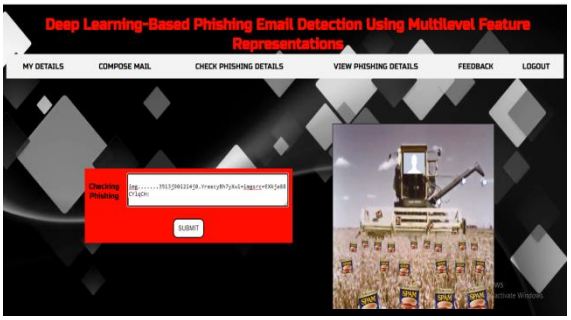


Fig 12: Compose mail module.

Description: This above page checking phishing details we give some links to check weather it is good message or phishing attack.

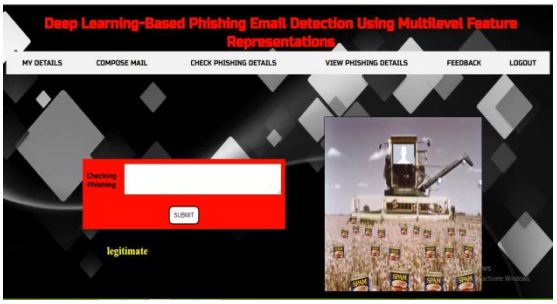


Fig 13: Check phishing details module.

Description: The checking phishing details we check some links and submit and it give legitimate that mean it is normal message.

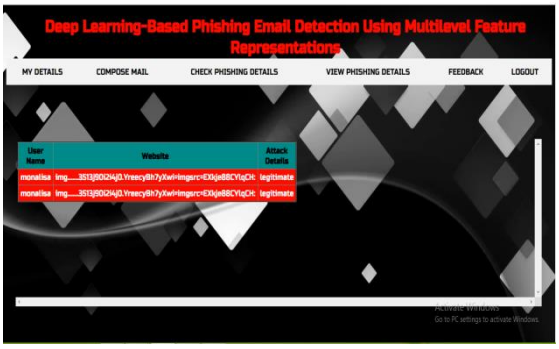


Fig 14: View phishing details module.

Description: The view phishing details we check some link that link are above the figure.



Fig 15: Feedback module.

Description: In the feedback we give message it is good message.

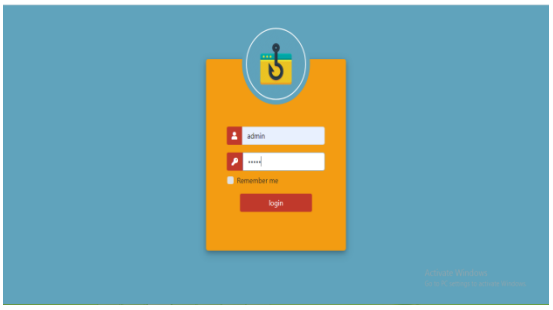


Fig 16: Admin module.

Description: In the above that details are store in admin site. We login the

admin side user and password, login.



Fig 17: User details module.

Description: In the user details mean in the user site we register the details. that details are shown in admin site.

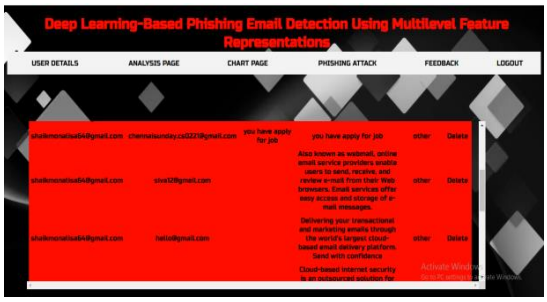


Fig 18: Analysis page module.

Description: In this analysis page sender can send to receiver what message send to receiver the message is social working or other message.

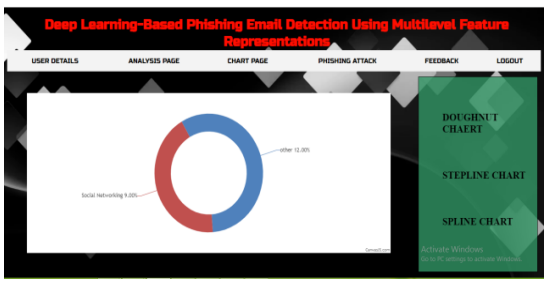


Fig 19: Chart page module.

Description: In this above chart page it checks the percentage of social networking and other message.

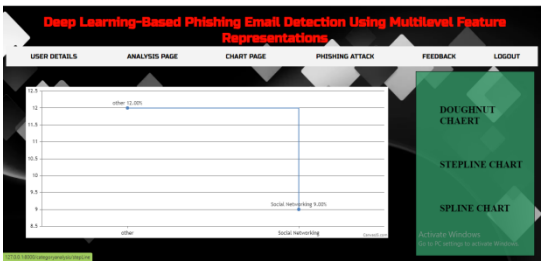


Fig 20: Chart page module.

Description: In this chart page to check the bar graph step line chart.

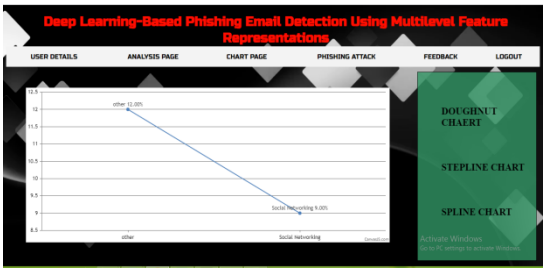


Fig 21: Chart page module.

Description: In this chart page to check the spline chart how much percentage social and other.

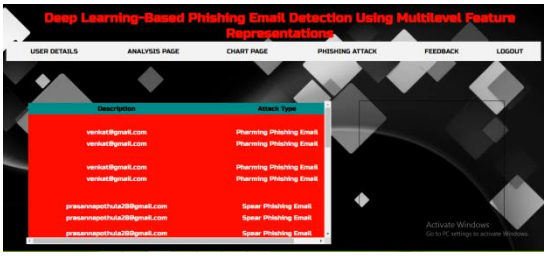


Fig 22: Phishing attack module.

Description: In this phishing attack has types in this it checks the message which types of message.

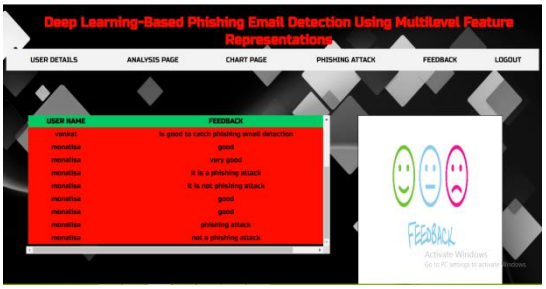


Fig 23: Feedback module.

Description: In this user site we give the feedback that feedback will show in admin site.



Fig 24: Logout module.

Description: In this admin site we logout the site.

Conclusion:

The Phishing Email Detection System using Improved RCNN with Multilevel Vector Attention Mechanisms is a technically feasible, cost-effective, and socially beneficial solution for combating phishing attacks. The RCNN model with attention mechanisms enhances detection accuracy by extracting key patterns in email content, ensuring real-time, high-precision phishing identification.

The system is scalable, integrates with enterprise email services, and complies with security regulations (GDPR, CCPA, HIPAA). Challenges like false positives and evasion techniques are addressed through model fine-tuning and adversarial training. The solution is highly effective, practical, and beneficial for individuals, businesses, and government organizations, helping to reduce cyber threats and enhance email security.

References:

[1] Anti-Phishing Working Group. (2018). Phishing Activity Trends Report1st Quarter2018 [Online]. Available:http://docs.apwg.org/reports/apwg_trends_report_q1_2018.

[2] PhishLabs. (2018). 2018 PhishTrends&IntelligenceReport.[Online].Available:https://info.phishlabs.com/hubfs/2018%20PTI%20Report/PhishLabs%20Trend%20Report_2018-digital.

[3] M. Nguyen, T. Nguyen, and T. H. Nguyen. (2018). ‘‘A deep learning model with hierarchical LSTMs and supervised attention for anti-phishing.’’ [Online].

[4] Anti Phishing Working Group. (2016).Phishing Activity Trends

Report4thQuarter2016.[Online]. Available:http://docs.apwg.org/reports/apwg_trends_report_q4_2016.

[5] Anti Phishing Working Group. (2015).PhishingActivityTrendsReport1st3rdQuarter2015.[Online].Available:http://docs.apwg.org/Preports/apwg_trends_report_q1-q3_2015.

[6] L. M. Form, K. L. Chiew, S. N. Sze, and W. K. Tiong, ‘‘Phishing email detection technique by using hybrid features,’’ in Proc. 9th Int. Conf. IT Asia (CITA), Aug. 2015, pp. 1–5.

[7] Microsoft. (2018). Microsoft Security Intelligence Report. [Online]. Available:https://clouddamcdnprodep.azureedge.net/gdc/gdcVAOQd7/original

[8] M. Hiransha, N. A. Unnithan, R. Vinayakumar, and K. Soman, ‘‘Deep learning based phishing e-mail detection,’’ in Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur. Privacy Anal. (IWSPA), A. D

[9] C. Coyotes, V. S. Mohan, J. Naveen, R. Vinayakumar, and K. P. Soman,‘‘ARES: Automatic rogue email spotter,’’ in Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur. Privacy Anal. (IWSPA), A. D. R. Verma, Ed. Tempe, AZ, USA, Mar. 2018

[10] S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, and C. Zhang, ‘‘An empirical analysis of phishing blacklists,’’ in Proc. 6th Conf. Email Anti-

[11] Spam(CEAS), Sacramento, CA, USA, 2009, pp. 1–10.Verma and N. Hossain, ‘‘Semantic feature selection for text with application to phishing email detection,’’ in Proc. Int. Conf. Inf. Secur. Cryptol. Cham, Switzerland: Springer, 2013, pp. 455–468,[12]G. Park and J. M. Taylor. (2015). ‘‘Using syntactic features for phishing detection.’’ [Online].

Available:

https://arxiv.org/abs/1506.00037

Author profile:



Mr. SK. Himam basha ,is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his Master

of Computer Applications (MCA) from Anna University, Chennai. With a strong research background, He has authored and co-authored research papers published in reputed peer-reviewed journals. His research interests include Machine Learning, Artificial Intelligence, Cloud Computing, and Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

.



Ms. P.Deepthi prasanna has received her MCA Masters of Computer Applications from QIS College of Engineering and Technology. (Autonomous), venga mukkapalem(V), Ongole,prakasam dist., Andhra Pradesh- 523272 affiliated to JNTUK in 2023-2025 .