DOI: https://doi.org/10.36719/2663-4619/116/42-45

Səyyad İbrahimov
Azərbaycan Texniki Universiteti
https://orcid.org/0009-0005-9224-6853
ibrahimovsayyad01@gmail.com

Fişinq hücumlarının aşkarlanmasında klassik mətn təsnifat metodlarının tətbiqi

Xülasə

Fişinq hücumları kibertəhlükəsizlik sahəsində ən geniş yayılmış və təhlükəli sosial mühəndislik üsullarından biri olaraq qalır. Bu hücumlar istifadəçiləri aldatmaqla onların şəxsi məlumatlarını ələ keçirməyə yönəlib. Bu problemin qarşısını almaq üçün effektiv və avtomatlaşdırılmış aşkarlama metodlarına ehtiyac vardır. Bu tədqiqatda klassik mətn təsnifatı yanaşmalarının fişinq mesajlarının aşkarlanmasında tətbiqi araşdırılır. Tədqiqat çərçivəsində Naive Bayes, Support Vector Machines (SVM), Decision Tree və Logistic Regression kimi ənənəvi təsnifat modellərinin fişinq və qeyri-fişinq mətnləri ayırd etmə qabiliyyəti qiymətləndirilmişdir. Bu metodların tətbiqi üçün əvvəlcə mətnlərdən xüsusiyyətlərin çıxarılması (TF-IDF, n-gramlar və s.) və məlumatların təmizlənməsi mərhələləri həyata keçirilmişdir. Modellərin effektivliyi dəqiqlik, həssaslıq, xatırlama və F1 dəyəri kimi qiymətləndirmə meyarları əsasında təhlil olunmuşdur.

Açar sözlər: *fişinq, kibertəhlükəsizlik, mətn təsnifatı, avtomatlaşdırılmış aşkarlama, süni intellektin tətbiqi*

Sayyad Ibrahimov
Azerbaijan Technical University
https://orcid.org/0009-0005-9224-6853
ibrahimovsayyad01@gmail.com

Application of Classical Text Classification Methods in Detecting Phishing Attacks

Abstract

Phishing attacks remain one of the most widespread and dangerous social engineering techniques in the field of cybersecurity. These attacks aim to deceive users and obtain their personal information. To prevent such threats, there is a growing need for effective and automated detection methods. This study investigates the application of classical text classification approaches in detecting phishing messages. Within the scope of the research, traditional classification models such as Naive Bayes, Support Vector Machines (SVM), Decision Tree, and Logistic Regression were evaluated for their ability to distinguish between phishing and non-phishing texts. Prior to the application of these models, feature extraction techniques (TF-IDF, n-grams, etc.) and data preprocessing steps were carried out. The performance of the models was analyzed based on evaluation metrics such as accuracy, precision, recall, and F1-score.

Keywords: *phishing, cybersecurity, text classification, automated detection, artificial intelligence applications*

Giriş

Rəqəmsal kommunikasiya texnologiyalarının sürətlə inkişaf etməsi ilə paralel olaraq kibertəhlükələr də dinamik şəkildə dəyişməkdədir. Bu təhdidlər arasında fişinq hücumları xüsusi yer tutur. Fişinq – istifadəçiləri aldatmaq yolu ilə onların şəxsi və maliyyə məlumatlarını əldə etməyə çalışan sosial mühəndislik texnikasıdır və bu hücum forması əsasən e-poçt, mesajlaşma platformaları

və saxta veb səhifələr vasitəsilə həyata keçirilir. İstifadəçi diqqətsizliyi və informasiya təhlükəsizliyi sahəsindəki boşluqlar bu hücumların effektivliyini artıran əsas amillərdir (Basnet, Sung, Liu, 2014).

Fişinq hücumlarının qarşısının alınması üçün ənənəvi müdafiə vasitələri (antivirus, firewall və s.) təkbaşına kifayət etmir. Bu səbəbdən son illərdə fişinq məzmunlarının avtomatik analiz və identifikasiyası üçün mətn təsnifatı metodlarından geniş istifadə olunur. Mətn təsnifatı – mətn əsaslı məlumatların əvvəlcədən təyin olunmuş kateqoriyalara ayrılmasını təmin edən süni intellekt və maşın öyrənməsi texnikalarının bir hissəsidir. Bu yanaşma, fişinq mesajlarının sintaktik və semantik xüsusiyyətlərini öyrənməklə onların digər mətnlərdən fərqləndirilməsinə imkan yaradır. Tədqiqatda klassik mətn təsnifatı metodlarının – Naive Bayes, Logistic Regression, Decision Tree və Support Vector Machines kimi modellərin – fişinq hücumlarının aşkarlanmasında effektivliyi təhlil olunur. Bu metodlar nisbətən sadə strukturları və aşağı hesablama tələbləri ilə seçilir, bu isə onları real vaxt rejimində çalışan təhlükəsizlik sistemləri üçün əlverişli edir. Həmçinin bu modellər interpretasiya imkanları baxımından daha şəffaf olduqları üçün praktiki tətbiqdə istifadəçi inamını artırır (Sahingoz, Buber, Demir, Diri, 2019, s. 345).

Tədqiqat

Bu araşdırmanın əsas məqsədi fişinq mətnlərinin aşkarlanmasında klassik metodların rolunu müəyyənləşdirmək, onların üstün və zəif cəhətlərini araşdırmaq və gələcəkdə daha kompleks modellərlə birgə istifadəsi üçün əsas yaratmaqdır. Fişinq hücumlarının aşkarlanmasında klassik mətn təsnifatı modellərinin tətbiqi həm nəzəri, həm də praktiki baxımdan mühüm əhəmiyyət kəsb edir. Fişinq mesajlarının əsas xüsusiyyəti onlarda insanı manipulyasiya edən və etibar doğuran ifadələrin tez-tez istifadə olunmasıdır. Bu cür mesajlar çox zaman rəsmi qurumların adından göndərilən, təcili reaksiya tələb edən və şəxsi məlumat istənilən müraciətlər şəklində olur. Bu cəhətlər onları adi reklam və ya məlumatlandırıcı mesajlardan fərqləndirir və mətn səviyyəsində analizlə aşkar edilə bilən xüsusi lingvistik nümunələrin formalaşmasına səbəb olur (Jain, Gupta, 2018).

Cədvəl 1. Klassik mətn təsnifatı modellərinin xüsusiyyətləri.

Model	İş Prinsipi	Sürət	Aydınlıq	Performans (orta)	Əsas Üstünlük
Naive Bayes	Ehtimal əsaslı təsnifat	Yüksək	Orta	Orta–Yüksək	Sadəlik və sürətli nəticə
Logisti Regression	Xətti ehtimallaşdırma	Orta	Yüksək	Yüksək	Aydın interpretasiya
Decision Tree	Qərar qaydaları və budaqlanma ilə təsnifat	Orta	Yüksək	Orta–Yüksək	Qaydaların vizuallaşdırılması
SVM	Hiper-səthlə verilənlərin ayrılması	Aşağı	Orta	Yüksək	Yüksək dəqiqlik

Mənbə: Sahingoz, Dogdu, 2019, s. 345–357).

Cədvəl 1-ə əsasən klassik mətn təsnifatı modellərinin əsas xüsusiyyətləri müqayisəli şəkildə təqdim olunub. Burada hər bir modelin necə işlədiyi, sürəti, nəticələrin interpretasiya imkanları, ortalama performans səviyyəsi və əsas üstünlükləri əks etdirilib. Naive Bayes modelinin sürətli və sadə olması, Logistic Regression-un yüksək dəqiqlik və aydın izah imkanı verməsi, Decision Tree-nin vizual interpretasiya üstünlüyü və SVM-in yüksək performansı önə çıxarılıb (Kowsari, Meimandi, Heidarysafa, Mendu, Barnes, Brown, 2019).

Cədvəl 2. Modellərin fişinq aşkarlanmasında performans göstəriciləri (Nümunəvi Dataset əsasında).

Model	Dəqiqlik (Accuracy)	Xatırlama (Recall)	Həssaslıq (Precision)	F1-Score
Naive Bayes	91.2%	88.5%	92.8%	90.6%
Logistic Regression	94.5%	93.1%	95.6%	94.3%
Decision Tree	89.7%	87.3%	90.2%	88.7%
SVM	96.3%	94.8%	97.2%	96.0%

Mənbə: (Verma, Hossain, 2020).

Cədvəl 2-yə əsasən fişinq aşkarlanması üçün istifadə olunan klassik modellərin konkret performans göstəriciləri – dəqiqlik (accuracy), xatırlama (recall), həssaslıq (precision) və F1-score – üzrə nəticələri verilmişdir. Nəticələrə əsasən SVM ən yüksək performansı göstərir, Logistic Regression onu izləyir. Naive Bayes və Decision Tree isə daha balanslı, lakin nisbətən aşağı göstəricilər təqdim edir. Bu məlumatlar modellərin tətbiq sahəsinə uyğun seçilməsini asanlaşdırır (Mohammadi, Hatzinakos, 2017).

Cədvəl 3. Mətn təsnifatı prosesinin mərhələləri.

Mərhələ	Açıqlama
Məlumatların toplanması	Fişinq və qeyri-fişinq mesajlarının toplanması
Ön emal (Preprocessing)	Kiçik hərflərə çevirmə, stop-word silinməsi, lemmatizasiya
Xüsusiyyət çıxarılması	TF-IDF, unigram/bigram, n-gramlar
Modelin qurulması	Naive Bayes, SVM və s. modellərin tətbiqi
Modellərin qiymətləndirilməsi	Dəqiqlik, F1-Score, ROC AUC və digər metriklərlə analiz
Real tətbiq	E-poçt sistemlərinə və təhlükəsizlik modullarına inteqrasiya

Mənbə: (Mohammadi, Hatzinakos, 2017, pp. 300–304).

Cədvəl 3-ə əsasən fişinq mesajlarının aşkarlanması məqsədilə tətbiq edilən mətn təsnifatı prosesinin mərhələləri göstərilmişdir. Buraya məlumatların toplanması, təmizlənməsi, xüsusiyyətlərin çıxarılması, model qurulması və qiymətləndirilməsi, eləcə də sistemin real tətbiqi daxildir. Bu mərhələlər təsnifat sisteminin düzgün və effektiv qurulması üçün əsas ardıcıllığı əks etdirir (Verma, Hossain, 2020).

Təhlil

Fişinq hücumları informasiya təhlükəsizliyi baxımından ciddi təhlükə yaradan, məqsədyönlü və aldadıcı sosial mühəndislik metodlarına əsaslanan hücumlardır. Bu hücumların əsas xüsusiyyəti – qanuni qurumların adından göndərilmiş kimi görünən mesajlar vasitəsilə istifadəçiləri məxfi məlumatlarını paylaşmağa məcbur etməkdir. Tədqiqatlar göstərir ki, istifadəçilərin əksəriyyəti bu tip mesajları ayırd etməkdə çətinlik çəkir və bu, fişinqin effektivliyini artırır. Bu problemin texnoloji həlli istiqamətində son illərdə süni intellektə əsaslanan avtomatlaşdırılmış sistemlər ön plana çıxıb. Bu sistemlərin təməlini isə mətn təsnifatı modelləri təşkil edir. Klassik mətn təsnifatı metodlarının əsas funksiyası – təqdim olunan mətnin hansı kateqoriyaya (bu halda fişinq və ya qeyri-fişinq) aid olduğunu müəyyənləşdirməkdir. Bu proses statistik analiz, xüsusiyyət çıxarılması və verilənlərin təlim vasitəsilə öyrənilməsi üzərində qurulur (Abu-Nimeh, Nappa, Wang, Nair, 2007).

Nəticə

Aparılan tədqiqat və analizlər göstərdi ki, klassik mətn təsnifatı modelləri – xüsusilə Naive Bayes, Logistic Regression, Decision Tree və Support Vector Machines – fişinq hücumlarının aşkarlan-

masında praktik və effektiv yanaşmalardan biri ola bilər. Bu modellər, düzgün strukturlaşdırılmış məlumat bazası və uyğun ön emal mərhələləri ilə birlikdə tətbiq edildikdə, yüksək dəqiqlik və qənaətbəxş nəticələr təmin edə bilir. Təcrübə nəticələri sübut edir ki, xüsusilə balanslı və təmizlənmiş mətn məlumatları ilə öyrədilmiş klassik modellər ilkin aşkarlama sistemləri üçün səmərəli seçimdir. SVM və Logistic Regression modelləri yüksək performans göstərsə də, Naive Bayes sadəliyi və sürəti ilə diqqət çəkir. Decision Tree isə interpretasiya baxımından daha əlverişli olub, təhlükəsizlik analitikləri üçün aydın qərar məntiqi təqdim edir.

Eyni zamanda müəyyən olunmuşdur ki, bu modellərin effektivliyi yalnız tətbiq edilən alqoritmlə deyil, həm də istifadə olunan xüsusiyyət çıxarma texnikaları, məlumatların emal səviyyəsi və təlim strategiyası ilə birbaşa əlaqəlidir. Bu amillər nəzərə alınmadan qurulmuş sistemlər gözlənilən nəticəni verməyə bilər. Nəticə etibarilə, klassik mətn təsnifatı metodlarının fişinq aşkarlama sistemlərində tətbiqi – aşağı resurslu mühitlərdə etibarlı bir alternativ olaraq çıxış edə bilər. Bununla yanaşı, gələcək tədqiqatlar bu metodların dərin öyrənmə modelləri ilə hibrid şəkildə birləşdirilməsi və adaptiv sistemlərin yaradılması üzərində fokuslanmalıdır. Bu istiqamət fişinq hücumlarına qarşı daha çevik və davamlı müdafiə sistemlərinin formalaşmasına imkan verəcək.

Ədəbiyyat

1. Basnet, R., Sung, A. H., & Liu, Q. (2014). Rule-based phishing attack detection. *Computers & Security, 39*, 17–33. https://doi.org/10.1016/j.cose.2013.04.012
2. Jain, A. K., & Gupta, B. B. (2018). Phishing detection: Analysis of visual similarity-based approaches. *Security and Privacy, 1*(1), e9. https://doi.org/10.1002/spy2.9
3. Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L. E., & Brown, D. F. (2019). Text classification algorithms: A survey. *Information, 10*(4), 150. https://doi.org/10.3390/info10040150
4. Mohammadi, M., & Hatzinakos, D. (2017). Phishing email detection using hybrid features and random forest classifier. In *2017 IEEE Conference on Communications and Network Security (CNS)* (pp. 300–304). IEEE. https://doi.org/10.1109/CNS.2017.8228655
5. Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from URLs. *Expert Systems with Applications, 117*, 345–357. https://doi.org/10.1016/j.eswa.2018.09.029
6. Verma, R., & Hossain, N. (2020). Natural language processing techniques for detecting phishing attacks: A literature survey. *Information Processing & Management, 57*(1), 102034. https://doi.org/10.1016/j.ipm.2019.102034
7. Ahmed, F., & Abulaish, M. (2013). A generic statistical approach for spam detection in social networking sites. *Computer Communications, 36*(10–11), 1120–1129. https://doi.org/10.1016/j.comcom.2013.03.004
8. Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007). A comparison of machine learning techniques for phishing detection. In *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit.* ACM, 60–69. https://doi.org/10.1145/1299015.1299021