9 maja 2024

AGH

Akademia Górniczo-Hutnicza
im. Stanisława Staszica
w Krakowie

Sekcja Teleinformatyki i Cyberbezpieczeństwa

61. Hutnicza Konferencja Studenckich Kół Naukowych AGH

Studenckie Koła Naukowe AGH

Akademia Górniczo-Hutnicza
im. Stanisława Staszica
w Krakowie

AGH

Studenckie Koła Naukowe AGH

MODEL SZTUCZNEJ INTELIGENCJI ROZPOZNAJĄCY POTENCJALNY PHISHING NA PODSTAWIE ANALIZY SKŁADNI HIPERŁĄCZA

Hubert Mąka

Koło Naukowe Telephoners

Wydział Informatyki, Elektroniki i Telekomunikacji AGH

Dr hab. inż. Marek Natkaniec

CZYM JEST PHISHING?

Według gov.pl:

"Phishing to jeden z najpopularniejszych typów ataków opartych o wiadomości e-mail lub SMS. Wykorzystuje inżynierię społeczną, czyli technikę polegającą na tym, że przestępcy internetowi próbują Cię oszukać i spowodować, abyś podjął działanie zgodnie z ich zamierzeniami."

"Wiadomości phishingowe są tak przygotowywane przez cyberprzestępców, aby wyglądały na autentyczne, ale w rzeczywistości są fałszywe. *Mogą próbować skłonić Cię do ujawnienia poufnych informacji, zawierać link do strony internetowej rozprzestrzeniającej szkodliwe oprogramowanie."*

Źródło: https://nordvpn.com/pl/blog/co-to-jest-phishing/

PRZECIEŻ TO TYLKO WIADOMOŚĆ CO W NIEJ NIEBEZPIECZNEGO?

Trochę danych:

*Według nowego raportu Hornetsecurity, który przeanalizował **45 miliardów** e-maili wysłanych w 2023 r., **phishing pozostaje metodą nr 1 cyberataków** stosowaną przez cyberprzestępców.*

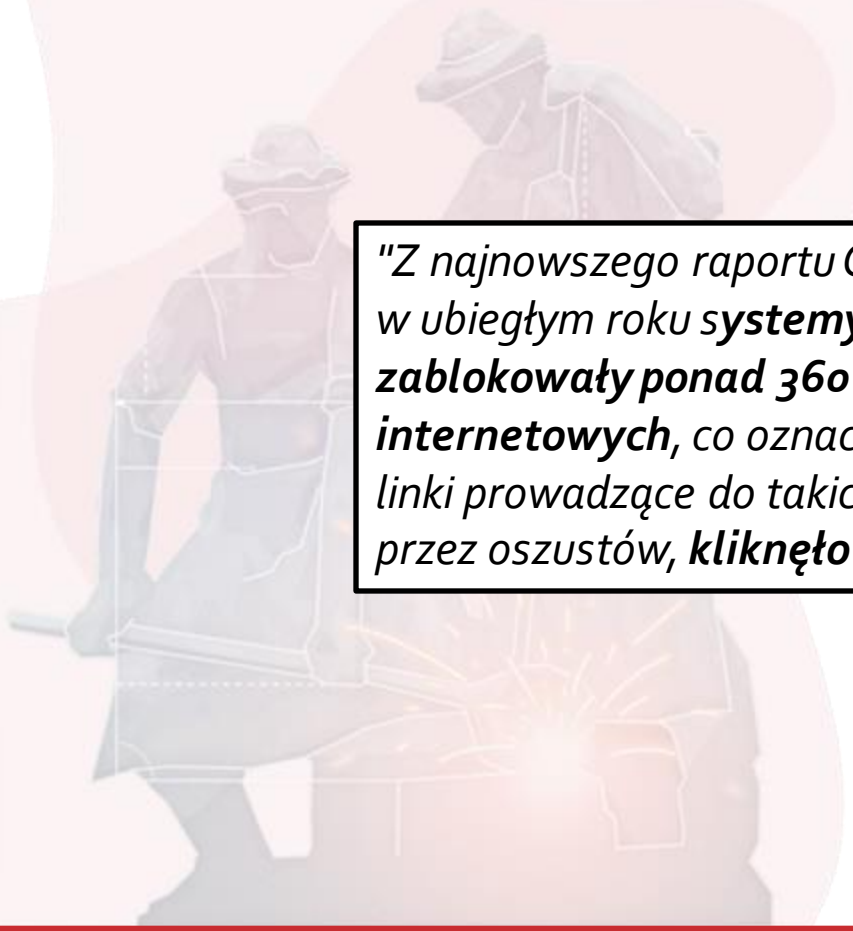
*"Phishing był najczęstszą metodą ataku za pośrednictwem poczty e-mail, **stanowiącą 43,3% ataków**"*

*„W przypadku tych e-maili najczęściej stosowaną techniką były złośliwe adresy URL – **30,5%** (wzrost o **144%** w porównaniu z rokiem ubiegłym). "*

***"3,6% wszystkich e-maili uznano za złośliwe.** Na pierwszy rzut oka wydaje się to „dobrą wiadomością"*

*"Ale jeśli weźmie się pod uwagę, że wciąż mówimy o **1,6 miliarda e-maili**, które narażają organizacje na ryzyko, to w rzeczywistości jest to straszna wiadomość"*

MOŻE W POLSCE JEST BEZPIECZNIEJ?



*"Z najnowszego raportu CERT Orange Polska wynika, że w ubiegłym roku s**ystemy bezpieczeństwa operatora zablokowały ponad 360 tys. fałszywych stron internetowych**, co oznacza aż trzykrotny wzrost r/r. W linki prowadzące do takich stron, stworzonych przez oszustów, **kliknęło ok. 5,5 mln Polaków.**"*

CZŁOWIEK - NAJSŁABSZE OGNIWO

Phishing jest wycelowany w największą słabość każego systemu - **człowieka**

"Najsłabszy element systemu definiuje jego bezpieczeństwo."

JAK WYKRYWAĆ?

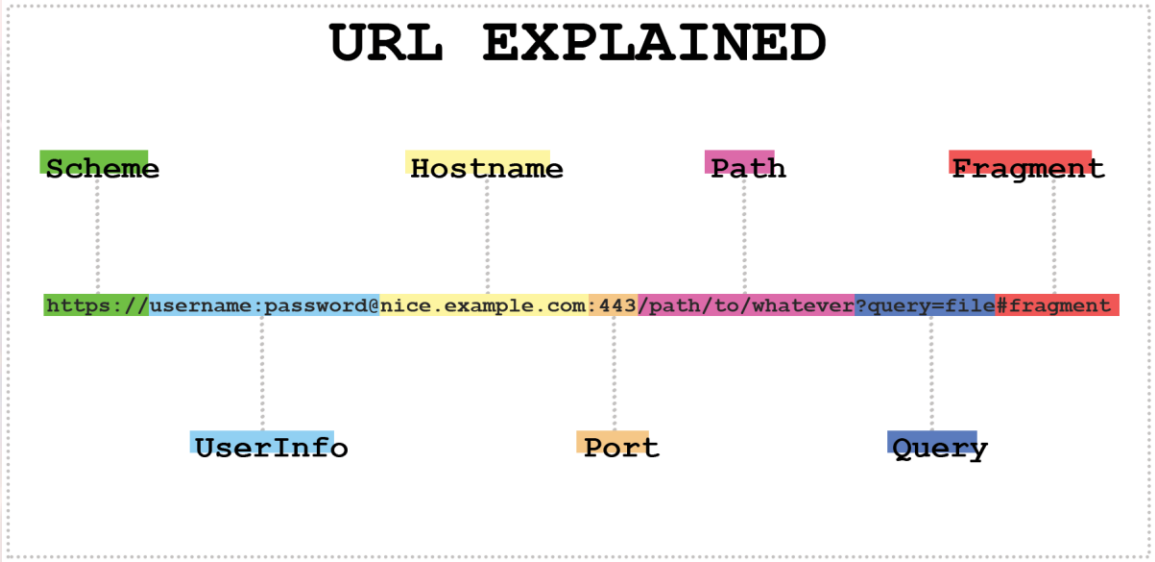
- **Metody Heurystyczne**
- DNS-Based (nie tylko) **Blacklist** (Google Safe Browsing API)
- **White-List**
- **Rankingi**
- **Uczenie Maszynowe**
- **Hybrydowe**

SKĄD DANE?

- **ISCX-URL2016** – University of New Brunswick
- **PhishTank**
- **Rankingi Najpopularniejszych Stron**

WYODRĘBNIONE CECHY:

- **Długość** URL
- Adres IP
- Czy dłuższy od średniej
- Występowanie **znaków**
- Czy użyty protokół to HTTPS
- Pewne **odchylenia** od typowego schematu URL
- Liczba **cyfr** w URL
- Liczba **liter** w URL
- **Głębokość** ścieżki plików
- Ilość kropek w domenie (głębokość domeny)
- **Długość** domeny
- URL skrócony
- Posiadanie prefixu **www.**
- Liczba słów w domenie



Źródło: https://ittavern.com/url-explained-the-fundamentals/

WYBRANE METODY KLASYFIKACJI:

	RF	ExtT	LR	SVM	MLP	KNN	GB
Acc	91.33%	89.68%	83.22%	82.20%	90.42%	90.95%	65.11%
Prec*	93.45%	94.99%	93.84%	95.20%	93.11%	93.66%	96.70%
Rec*	88.59%	83.43%	70.53%	67.24%	86.97%	87.54%	30.16%

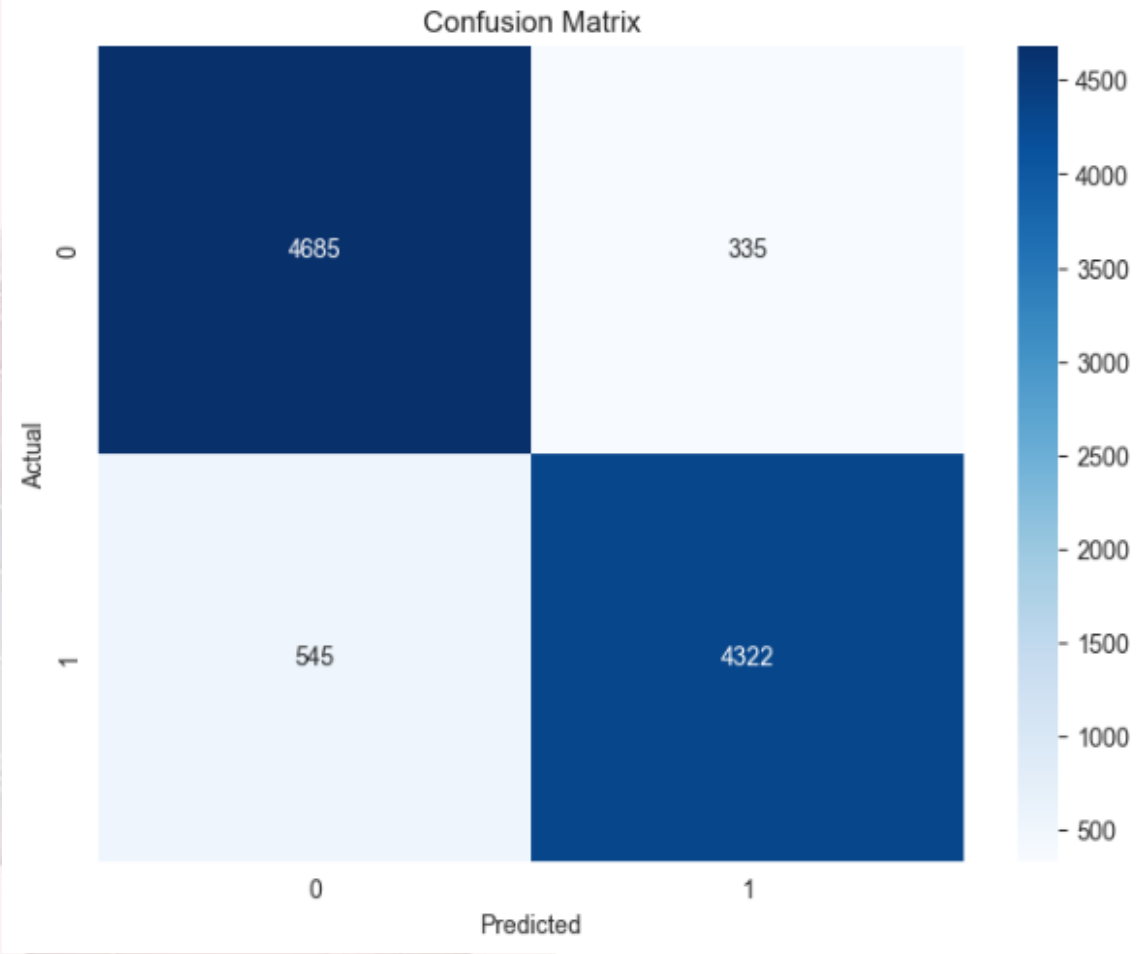
- Random Forest
- Extra Trees
- Logistic Regression
- Support Vector Machines
- Multilayer Perceptron
- K-Nearest Neighbors
- Gradient Boosting

*Prec i Rec obliczone dla klasy 1

PARAMETRY MODELU RF:

- **Dla random state: 41**
- **Liczba drzew: 21**
- **Maksymalna głębokość: 20**

TABLICA POMYŁEK RF:



Akademia Górniczo-Hutnicza
im. Stanisława Staszica
w Krakowie

AGH

Studenckie Koła Naukowe AGH

PREZENTACJA DZIAŁANIA MODELU

ŹRÓDŁA:

- Buber, E., Demir, Ö., & Sahingoz, O. K. (2017). Feature selections for the machine learning based detection of phishing websites. In 2017 International Artificial Intelligence and Data Processing Symposium (IDAP). IEEE.
- Mohammad, R. M., Thabtah, F., & McCluskey, L. (2012). An assessment of features related to phishing websites using an automated technique. In 2012 International Conference. IEEE.
- Reyes-Dorta, N., Caballero-Gil, P. & Rosa-Remedios, C. Detection of malicious URLs using machine learning. Wireless Netw (2024). Springer. https://doi.org/10.1007/s11276-024-03700-w
- https://www.hornetsecurity.com/en/cyber-security-report/
- https://cert.orange.pl/

Akademia Górniczo-Hutnicza
im. Stanisława Staszica
w Krakowie

AGH

Studenckie Koła Naukowe

AGH

DZIĘKUJĘ ZA UWAGĘ