УДК 004.056:004.912

ЗАСТОСУВАННЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТІВ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ. ЧАСТИНА 1. ЗАХИСТ ВІД ФІШИНГУ

Панков Д.С., Машура А.П.
e-mail: andrii.mashura@nure.ua
Науковий керівник - к.т.н., доц. Горелов Д.Ю.
Харківський національний університет радіоелектроніки, каф. КРіСТЗІ
студ. наук. гурток «Біометричні технології контролю доступу»
м. Харків, Україна

An algorithm for building a spam filter to detect phishing attacks is proposed. The accuracy of classification of emails from the "Shuhaib-Ahamed Dataset" was studied in the Orange software environment. The accuracy of detecting texts containing phishing features was 98.5 % when using the fastText extractor and a neural network as a classifier. The proposed algorithm can be used in addition to the spam/phishing filter built into the email client or as an add-on to a DLP system to prevent the spread of phishing messages through the corporate network due to staff negligence or insider activity.

Метою фішингу є отримання доступу до конфіденційних даних користувачів або встановлення шкідливого програмного забезпечення з використанням методів соціальної інженерії.

Головний інструмент фішингу – електронна пошта. На поштову скриньку жертви зловмисник надсилає листа, який має спонукати її ввести необхідну інформацію. Також повідомлення може містити шкідливе вкладення, яке при його активації проникне в систему, збиратиме і надсилатиме інформацію зловмиснику. Особливість класичного фішингу – масове розсилання листів з ідентичним змістом.

Цільовий фішинг (Spear-Phishing), на відміну від звичайного, спрямований на досягнення конкретної мети, а, відповідно, є набагато небезпечнішим, оскільки кіберзлочинці спеціально збирають інформацію про жертву, щоб зробити своє послання переконливішим. Якісно зроблений лист для цільового фішингу іноді дуже важко відрізнити від цілком легітимного листа, який не переслідує шкідливих цілей. Дані особливості зробили цей тип атак одним із найбільш ефективних.

Таким чином, однією зі складових комплексної системи з захисту від фішингових атак є спам-фільтр, задача якого – в реальному режимі часу проводити автоматичну класифікацію змісту електронних листів та блокувати ті з них, що мають ознаки спаму / фішингу.

В дослідженнях використовувався датасет «Shuhaib-Ahamed Dataset», що містить 32129 електронних листів, 16032 з яких є спам-листами / фішинговими листами, та 16097 – це звичайні електронні листи. Основні характеристики датасету наведені на рис. 1 – рис. 2.

В якості програмного засобу для проведення досліджень було використано інструмент для візуалізації даних, машинного навчання та інтелектуального аналізу даних Orange. В якості екстрактору інформативних ознак використовувались два алгоритми: нейронна мережа fastText від Facebook (перетворює текст на вектор з 300 ознак) та нейронна мережа SBERT від Google (перетворює текст на вектор з 384 ознак). В якості алгоритму класифікації користувачів використовувались методи Logistic Regression, SVM, Naive Bayes, Neural Network. Точність класифікації перевірялась за вбудованим у віджет «Test and Score» алгоритмом 5-fold cross-validation.

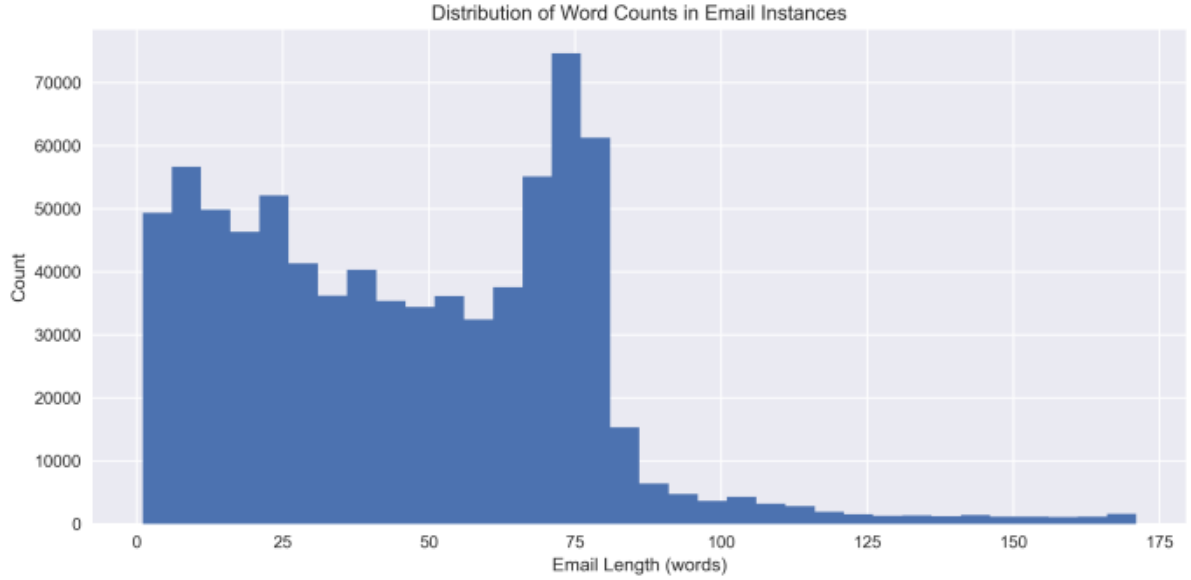

Рисунок 1 – Гістограма довжин (кількість слів) текстів
з датасету «Shuhaib-Ahamed Dataset»

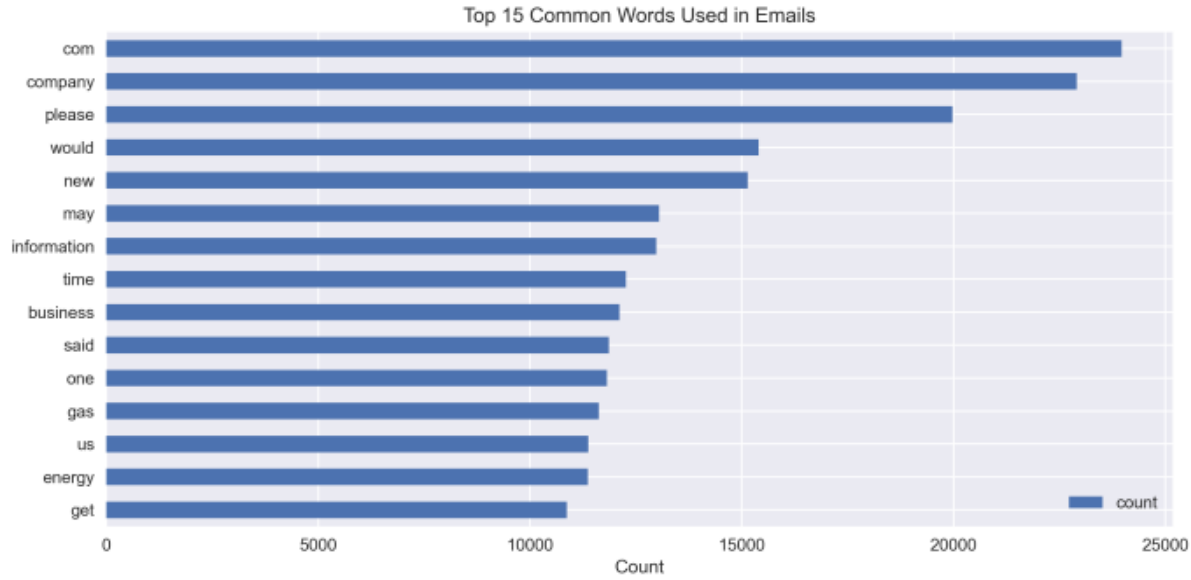

Рисунок 2 – Топ 15 найуживаніших слів в текстах
датасету «Shuhaib-Ahamed Dataset»

Результати класифікації текстів з датасету «Shuhaib-Ahamed Dataset» наведено на рис. 3 та рис. 4 для екстракторів fastText та SBERT відповідно. Як можна бачити, для обох варіантів екстрактору алгоритм класифікації Neural Network забезпечив максимальну точність, досягши значення 98.5 %. Дещо гірші показники показав метод Logistic Regression з максимальним значенням точності класифікації – 97.4 %.

Model	AUC	CA	F1	Prec	Recall
Logistic Regression	0.983	0.943	0.943	0.943	0.943
SVM	0.730	0.638	0.636	0.640	0.638
Naive Bayes	0.774	0.682	0.681	0.686	0.682
kNN	0.982	0.928	0.928	0.929	0.928
Neural Network	0.998	0.985	0.985	0.985	0.985

Рисунок 3 – Результати класифікації текстів з датасету «Shuhaib-Ahamed Dataset». Екстрактор – нейронна мережа fastText

Model	AUC	CA	F1	Prec	Recall
Logistic Regression	0.996	0.974	0.974	0.974	0.974
SVM	0.961	0.898	0.898	0.899	0.898
Naive Bayes	0.984	0.935	0.935	0.935	0.935
kNN	0.996	0.973	0.973	0.973	0.973
Neural Network	0.997	0.982	0.982	0.982	0.982

Рисунок 4 – Результати класифікації текстів з датасету «Shuhaib-Ahamed Dataset». Екстрактор – нейронна мережа SBERT

Таким чином, подібний агент можна використовувати на додаток до вбудованого в email клієнт спам / фішинг-фільтру або на додаток до DLP сис-теми з метою запобігання поширенню засобами корпоративної мережі фішингових повідомлень через недбалість персоналу або діяльність інсайдерів.

Список використаних джерел:

1. 22. Shuhaib-Ahamed Dataset. URL: https://github.com/Shuhaib-Ahamed/Email-Spam-Classification?tab=readme-ov-file (дата звернення: 01.12.2024)

2. Vasyl Alieksieiev, Aleksey Strelnitskiy, Dmitry Gavva, Denis Gorelov, Yuliia Synytsia. Studying of keystroke dynamics statistical properties for biometric user authentication. Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), Pages 559-563, 2018.

3. Дослідження інформативних параметрів диграфів клавіатурного почерку для задач ідентифікації користувачів комп'ютерних мереж / Д.Ю. Горелов, О.О. Іванова, О.В. Кокорін, Д.В. Маслій, О.В. Литвиненко // Радіотехніка: Всеукр. Міжвід. Наук.-техн. Зб. – 2020. – вип. 201. – с. 194 – 200.