UNIVERZITA PAVLA JOZEFA ŠAFÁRIKA
PRÍRODOVEDECKÁ FAKULTA

DETEKCIA FORIEM SOCIÁLNEHO INŽINIERSTVA V EMAILOVEJ KOMUNIKÁCII

2019 **Eva MARKOVÁ**

UNIVERZITA PAVLA JOZEFA ŠAFÁRIKA
PRÍRODOVEDECKÁ FAKULTA

DETEKCIA FORIEM SOCIÁLNEHO INŽINIERSTVA V EMAILOVEJ KOMUNIKÁCII

BAKALÁRSKA PRÁCA

Študijný program:	Informatika
Pracovisko (katedra/ústav):	Ústav informatiky
Vedúci bakalárskej práce:	Mgr. Tomáš Bajtoš

Košice 2019 **Eva MARKOVÁ**

Univerzita P. J. Šafárika v Košiciach
Prírodovedecká fakulta

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Eva Marková
Študijný program: Informatika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor: 9.2.1. informatika
Typ záverečnej práce: Bakalárska práca
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Detekcia foriem sociálneho inžinierstva v emailovej komunikácii

Názov EN: Detection of social engineering forms in email communication

Cieľ:
1) Analýza a spracovanie foriem sociálneho inžinierstva v emailovej komunikácii
2) Porovnanie a spracovanie aktuálnych prístupov k ochrane voči sociálnemu inžinierstvu v emailovej komunikácii
3) Návrh a implementácia systému na detekciu foriem sociálneho inžinierstva v emailovej komunikácii a jeho otestovanie

Literatúra:
1. MANN, Ian. Hacking the human: social engineering techniques and security countermeasures. Routledge, 2017.
2. BULLÉE, Jan#Willem Hendrik, et al. On the anatomy of social engineering attacks—A literature#based dissection of successful attacks. Journal of investigative psychology and offender profiling, 2018, 15.1: 20-45.
3. GUPTA, B. B., et al. Fighting against phishing attacks: state of the art and future challenges. Neural Computing and Applications, 2017, 28.12: 3629-3654.

Vedúci: Mgr. Tomáš Bajtoš
Ústav : ÚINF - Ústav informatiky
Riaditeľ ústavu: prof. RNDr. Viliam Geffert, DrSc.

Dátum schválenia: 03.05.2019

Univerzita Pavla Jozefa Šafárika v Košiciach
Prírodovedecká fakulta
Ústav informatiky

Poďakovanie

Týmto sa chcem poďakovať vedúcemu svojej práce Mgr. Tomášovi Bajtošovi za odborné vedenie, cenné rady a veľkú pomoc pri spracovaní problematiky sociálneho inžinierstva v emailovej komunikácii a celkovo počas tvorby práce.

Abstrakt v štátnom jazyku

Neoddeliteľnou súčasťou každodenného života a práce na počítači je vlastníctvo a využívanie emailovej adresy. Vo všeobecnosti sa človek považuje za najslabší článok bezpečnosti, a preto je jednou z najväčších hrozieb v dnešnom virtuálnom svete práve sociálne inžinierstvo. Hlavným cieľom tejto práce je analyzovať existujúce prístupy na detekciu sociálneho inžinierstva, porovnať ich, navrhnúť a implementovať systém, ktorý by bol schopný detegovať podvrhnuté emaily, a následne ich kategorizovať do 4 vybraných kategórií – spam, scam, phishing a legitímne emaily. Na detekciu podvrhnutých správ sme extrahovali z emailov 11 atribútov, a tiež sme sa pozreli na prítomnosť najfrekventovanejších slov pre každú kategóriu. Na zatriedenie emailu do kategórie využívame často používaný algoritmus – „Random Forest“, ktorý nadobúdal najlepšie výsledky pre klasifikáciu. Vytvorený dataset obsahuje 148 podvrhnutých a 150 legitímnych emailov.

Kľúčové slová: phishing, random forest, scam, sociálne inžinierstvo, spam.

Abstrakt v cudzom jazyku

An inherent part of everyday life and work on computer is ownership and use of an email address. In general, human is considered to be the weakest part in security, and therefore one of the greatest threats in today's virtual world is social engineering. The main goal of this work is to analyze existing approaches to social engineering detection, compare them and implement a system that would be able to detect illegitimate emails and then categorize them into 4 selected categories - spam, scam, phishing and legitimate emails. We extracted 11 attributes from emails to detect spoofed messages, and we also looked at the presence of the most frequent words for each category. To categorize email, we use the often used "Random Forest" algorithm to get the best results for classification. The created dataset contains 148 illegitimate and 150 legitimate emails.

Key words: phishing, random forest, scam, social engineering, spam.

Obsah

Zoznam ilustrácií

Zoznam tabuliek

Zoznam skratiek a značiek

CART	Classification and regression trees, algoritmy rozhodovacích stromov
CS-SVM	Cuckoo Search – Support Vector Machines, modifikovaná metóda podporných vektorov
CSIRT	Computer Security Incident Response Team – tím pre riešenie počítačových bezpečnostných incidentov
DNS	Domain Name System, systém mien domén
DT	Decision Tree, rozhodovací strom
FN	False Negative, „neskutočne negatívna“ hodnota pre výkonnostnú metriku
FP	False Positive, „neskutočne pozitívna“ hodnota pre výkonnostnú metriku
IP	Internet Protocol
k-NN	k-Nearest Neighbors, algoritmus k-najbližších susedov
LDA	Latent Dirichlet Allocation, latentné dirichletové rozdelenie
MLP	Multilayer Perceptron, viacvrstvový perceptrón
NLP	Natural Language Processing, spracovanie prirodzeného jazyka
RF	Random Forest, náhodný les
SVM	Support Vector Machines, metóda podporných vektorov
TN	True Negative, „skutočne negatívna“ hodnota pre výkonnostnú metriku
TP	True Positive, „skutočne pozitívna“ hodnota pre výkonnostnú metriku
URL	Uniform Resource Locator, jednotný vyhľadávač zdrojov

Úvod

V dnešnej dobe väčšina ľudí využíva či už v domácnosti, alebo v práci počítač, v rámci neho internet a na komunikáciu využívajú emailové schránky. Podľa štatistického portálu Statista sa od roku 2014 zvýšil počet aktívnych emailových adries o takmer 1 500 miliónov, v dnešnej dobe to teda predstavuje takmer 5 600 miliónov aktívnych emailových adries. [1]

Častokrát sa stáva, že užívateľ obdrží email, pri ktorom nevie určiť, či sa jedná o legitímny alebo podvrhnutý email. Je dôležité, aby každý užívateľ mal možnosť overiť si skutočnosť, či je jeho emailová komunikácia naozaj bezpečná.

Existuje mnoho rôznych riešení na detekciu podvodných emailov, čo ale chýba, je ich kategorizácia a vytvorenie komplexného systému, ktorý by korektne fungoval a naozaj informoval užívateľa o tom, či môže veriť tomu, čo si prečítal. V tejto práci sa zameriavame na kategorizáciu emailov hlavne s využitím algoritmov dolovania dát, keďže sa ukázalo, že sú najlepšou možnou voľbou. Práca je rozdelená do piatich kapitol.

V prvej kapitole sa venujeme sociálnemu inžinierstvu ako takému a jeho konkrétnym formám v emailovej komunikácii, medzi ktoré patrí spam, scam, phishing, spearphishing a hoax.

V rámci druhej kapitoly sledujeme rôzne prístupy na detekciu foriem sociálneho inžinierstva v emailovej komunikácii a porovnávame existujúce riešenia. Zameriavame sa hlavne na úspešnosť týchto prístupov.

Ďalšie dve kapitoly sa venujú praktickej časti práce – prípravnej fáze pre zatrieďovanie emailov a samotnej tvorbe systému. Tieto časti spolu veľmi úzko súvisia, keďže bez datasetu by sme nemohli vytvoriť korektne fungujúci systém na identifikáciu podvodných emailov.

V poslednej kapitole sme sa zamerali na vyhodnotenie systému, pričom dostal na vstup testovacie údaje a sledujeme, ako sa zachová – v našom prípade ako zaklasifikuje emaily do štyroch tried.

1 Sociálne inžinierstvo

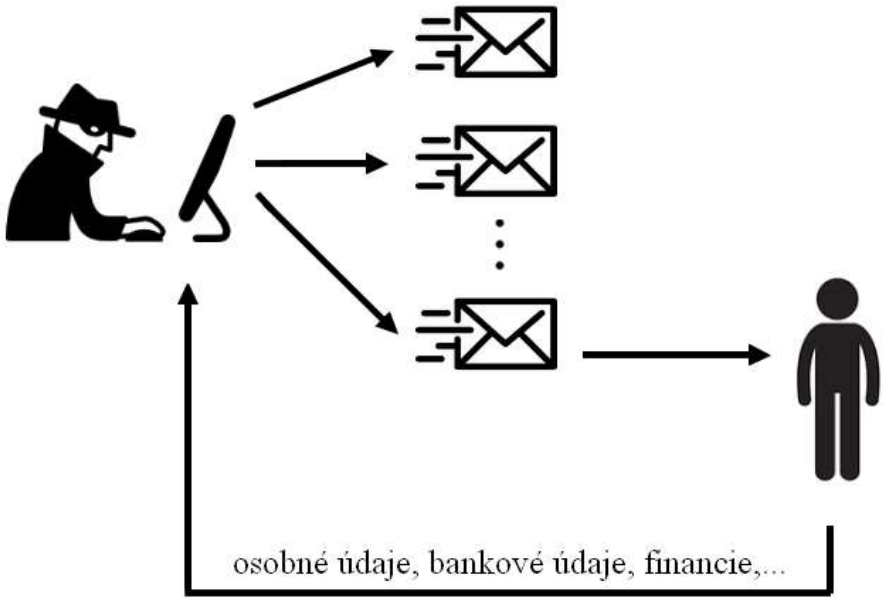
Sociálne inžinierstvo [2] je metóda získavania prístupu do systémov, k údajom alebo budovám skúmaním ľudskej psychológie. Namiesto používania technických zručností alebo fyzického násilia, sociálne inžinierstvo zahŕňa netechnické schémy, ktoré vyvíjajú útočníci. V deväťdesiatych rokoch Kevin Mitnick, známy hacker, popularizoval výraz „sociálne inžinierstvo“ aj napriek tomu, že pokusy využívajúce prvky sociálneho inžinierstva za účelom prezradenia citlivých údajov, poznáme od nepamäti. Útoky pomocou sociálneho inžinierstva sú zvyčajne konštruované outsidermi, ktorí používajú psychologické triky, aby im používateľ dal prístup k počítaču alebo údajom. Tieto útoky sú často vykonávané medzi zamestnancami jednej organizácie.

Uvádza sa, že len 30 percent útokov vykonávajú outsideri, ktorí nie sú súčasťou organizácie, ktorá je obeťou. [3] To znamená, že až 70 percent útočníkov pochádza priamo z organizácie. Toto je potrebné brať do úvahy, ak sa snažíme zabezpečiť firmu proti sociálnym inžinierom.

Ako už bolo spomenuté, cieľom sociálneho inžinierstva je prinútiť ľudí, aby útočníkom poskytli presne to, čo potrebujú. Sociálni inžinieri sa „živia“ tým, akí ľudia v skutočnosti sú:

- Ľudia častokrát chcú byť nápomocní, ale to môže viesť k úniku citlivých údajov.
- Pre ľudí je prirodzené veriť iným až kým im nedokážu, že sa im nedá veriť.
- Majú strach, že sa dostanú do problémov.
- Lepia si heslá na obrazovky, poprípade nechávajú dôležité dokumenty na mieste, kde k nemu majú prístup aj iní ľudia.

Proces sociálneho inžinierstva vo všeobecnosti vyzerá ako na obr. 1:



Obr. 1 Proces sociálneho inžinierstva

Útočník rozošle niekoľko emailov a čaká, či sa niektorá z obetí nechá oklamať a sprístupní mu napríklad osobné údaje, bankové údaje alebo odošle nejaké peniaze.

1.1 Typy sociálneho inžinierstva

Hoci sociálni inžinieri majú najväčší úspech s interakciou ľudí, niektoré počítačovo založené metódy im umožňujú získať požadované údaje použitím softvérov. Jedna z najlepších metód bola prvýkrát predstavená Internetu vo februári 1993. Užívatelia boli vyzvaní na zadanie prihlasovacích údajov do systému, avšak po zadaní správnych údajov sa im zobrazila výzva na vyplnenie znovu tých istých údajov. Útočníkovi sa podarilo nainštalovať program, kde užívatelia zadali údaje, zhromaždil informácie, a potom ich presmeroval na reálny portál. Podľa zverejnených článkov v tom čase až 95% bežných užívateľov poskytlo svoje údaje. [2]

Sociálne inžinierstvo môže byť rozdelené na 2 typy – **založené na ľuďoch** a **založené na technológii**. Sociálne inžinierstvo založené na ľuďoch odkazuje na interakciu medzi ľuďmi a sociálne inžinierstvo založené na technológii znamená mať nejaké elektronické rozhranie na získanie požadovaných údajov. [2]

1.1.1 Sociálne inžinierstvo založené na ľuďoch

V rámci tejto kategórie poznáme štyri menšie podkategórie. Prvou je „**napodobňovanie a dôležitý užívateľ**“, pričom sa útočník vydáva za inú entitu. Vydávanie sa za inú osobu, poprípade jednoduché „klebetenie“ sú typické príklady ako sociálni inžinieri získavajú informácie. Často kontaktujú v organizácii helpdesk pre zisk mien kľúčových zamestnancov. A pokiaľ majú to, čo potrebujú na získanie prístupu, zaútočia na zraniteľnejšiu osobu – na niekoho, kto má informáciu, ktorú hľadajú. [2]

Ďalšou podkategóriou je „**autorizácia tretej strany**“. Útočník sa odvoláva na vyššiu autoritu, aby získal prístup do systému, ku ktorému nemá povolenie. Je dôležité nezabudnúť na to, že najviac sociálnych inžinierov sú insideri (ľudia, ktorí sú súčasťou organizácie). [3]

„**In person**“ je ďalšia podkategória a hovorí o tom, že útočníci vstúpia do budovy predstierajúc, že sú zamestnanci, návštevníci alebo služobný personál. Zamestnanci nechávajú zapnuté zariadenia aj v prítomnosti služobného personálu, poprípade ich nechajú bez dozoru vo svojich kanceláriách. A tu vyvstáva problém, čo ak personál je len „prezlečený“ útočník?

Poslednou kategóriou je „**Dumpster Diving**“ a „**Shoulder Surfing**“. Jedná sa o dve najstaršie formy sociálneho inžinierstva. „Dumpster Divers“ sú ochotní „zašpiniť sa“, aby získali informácie, ktoré potrebujú. Organizácie častokrát vyhadzujú dôležité dokumenty, pričom citlivé údaje, manuály alebo napríklad zoznamy telefónnych čísel by mali byť pred vyhodením skartované, a to sa častokrát zanedbáva. „Shoulder Surfer“ pozerá cez rameno obete, aby získal napríklad heslo alebo pinkód k platobnej karte.

1.1.2 Sociálne inžinierstvo založené na technológii

V rámci tejto kategórie sú známe 3 rôzne podkategórie. Prvou z nich je „**Pop-Up Windows**“. Na obrazovke sa zjaví okno, ktoré informuje užívateľa napríklad o tom, že pripojenie k sieti bolo prerušené a sieťové pripojenie musí byť znovu autentifikované. Po vyplnení údajov ich pop-up program pošle útočníkovi. Ďalšou možnosťou je žiadať osobné údaje do systémov. Túto formu sociálneho inžinierstva nazývame phishing. [2]

Prílohy emailov môžu byť tiež veľmi nebezpečné. Škodlivé programy a spúšťacie súčasti sú častokrát ukryté v emailoch, práve preto je dôležité byť ostražitý pri sťahovaní akejkoľvek prílohy emailu.

Poslednou kategóriou sú **webové stránky**. Častokrát sa odosielajú emaily, ktoré sa vydávajú za legitímne, pričom cieľom je získať citlivé údaje a podobne. Vláda, finančné inštitúcie, online platby, aukcie a podobne sú častým cieľom útokov.

My sa v našej práci zameriavame práve na sociálne inžinierstvo založené na technológiách, pretože cieľom je detegovať podvodné emaily v rámci emailovej komunikácie a následne ich kategorizovať. Emaily zahŕňajú rôzne kombinácie všetkých troch podkategórií, pretože veľa podvodných emailov obsahuje práve škodlivé odkazy, prílohy alebo dokonca vyzývajú k vyplneniu osobných údajov.

1.2 Formy sociálneho inžinierstva v emailovej komunikácii

V rámci emailovej komunikácie rozlišujeme niekoľko foriem sociálneho inžinierstva. Spomenieme si tie najdôležitejšie:

1.2.1 Phishing

Phishing [4] je postup používaný na oklamanie užívateľov, aby poskytli svoje osobné alebo finančné údaje útočníkovi. Útoky tohto typu sú iniciované cez emaily, pričom obsahujú napríklad odkazy na škodlivé domény, ktoré sa na prvý pohľad zdajú byť legitímne. Útočníci sa snažia presvedčiť svoju obeť, aby navštívila podvrhnutú stránku, ktorá vyzerá identicky ako pôvodná legitímna stránka. Takéto stránky sú vytvorené za účelom stiahnutia malvéru do počítača, poprípade na získanie osobných údajov a podobne.

Procedúry a techniky phishingu sa vyvíjajú veľmi rýchlo. Útočníci často vedia ako funguje počítačová komunikácia, protokoly a podobne, preto je pre nich jednoduchšie prekonať bezpečnostné protokoly a teda zvyšujú úspešnosť útokov. Neschopnosť užívateľa detegovať phishingové útoky, zvyšovanie počtu útokov a diverzita týchto útokov spôsobuje nárast úspechu phishingu. Ako už bolo spomenuté phishing je nie len sociálne inžinierstvo, ale tiež technologický problém.

Phishing je jeden z útokov, kde útočník vystupuje ako niekto iný a v závislosti od reputácie a vzťahu s obeťou sa snaží odhaliť informácie.

1.2.2 Spearphishing

Spearphishing [5] je forma útoku, ktorá slúži na získanie prístupu do systému alebo organizácie, aby sa vykonal útok alebo špionáž. Útočníci pred samotným útokom zisťujú informácie o užívateľoch, aby mohli vytvoriť personalizované emailové správy, pričom sa vydávajú za dôveryhodné entity, známych a podobne. Aby bol útok úspešný, musí užívateľ tiež kliknúť na odkaz. Jedná sa o phishing, ktorý je osobnejší a preto má o niečo väčšiu mieru úspešnosti ako samotný phishing.

1.2.3 Spam

Spam [6] je irelevantná alebo nevyžiadaná správa rozposielaná cez Internet typicky veľkému množstvu užívateľov najčastejšie za účelom reklamy.

Spam sa vyznačuje tým, že spôsobuje zahltenie úložného priestoru emailovej schránky a tiež zneužívanie výpočtového výkonu. Spam spôsobuje, že ľudia musia triediť emaily a vďaka tomu nielenže plytvajú časom, ale sú z toho tiež podráždení. A čo je najhoršie, častokrát spam spôsobuje právne problémy reklamou pornografie, pyramídových schém a podobne.

Problém týchto nežiadúcich správ je dnes seriózny, pretože spam zahlcuje veľkú časť emailovej schránky a spamové filtre sú v mnohých prípadoch málo účinné.

1.2.4 Scam

Nigérijský **scam** [7] inak známy aj ako ‘419‘ scam je populárna forma podvodu, pri ktorej podvodník presvedča obeť, aby zaplatila určitú sumu peňazí na základe sľubu budúcej väčšej odmeny. Tento názov zahŕňa viacero variácií podvodov ako napríklad falošná lotéria, podvod s čiernymi peniazmi a podobne. Tento fenomén začínal pomocou klasickej pošty, neskôr sa začal rozposielať pomocou faxov a dnes sa s ním stretávame aj v emailovej komunikácii.

Ako výsledok nahlasovania podvodov takéhoto typu vznikli stránky, kde sa dajú takéto emaily nahlasovať, napríklad 419scam.org.

V dnešnej dobe vnímame scam ako jedinečný typ spamu. Zatiaľ čo väčšina spamu je rozosielaná hromadne pomocou botnetov a kompromitovaných strojov, tieto scamy sa vo veľkej miere stále píšu ručne. Scammeri sa spoliehajú na ľudský faktor: škoda, chamtivosť a techniky sociálneho inžinierstva. Používajú veľmi primitívne nástroje

v porovnaní s inými formami sociálneho inžinierstva, kde sú všetky operácie častokrát úplne automatizované. Aj napriek tomu, že tieto scamy sú tienené spamom, ktorý chodí v omnoho väčšej miere, stále predstavujú problém, ktorý spôsobuje stratu financií u množstva ľudí vo svete.

1.2.5 Hoax

Hoaxy [8] sú známe tým, že nie sú škodlivé pre akýkoľvek systém. Hoax sa považuje za nevyžiadanú poštu, ktorá môže používateľom alebo čitateľom emailu poskytnúť zavádzajúce informácie. Prenášajú klamlivé informácie, pričom ich predstavujú ako pravdivé. S postupom času hoaxy vyzerajú presvedčivejšie, a práve preto sa veľa ľudí nechá oklamať. To spôsobuje častokrát stratu financií a iritáciu individuálnych užívateľov. Čo je horšie, hoax má schopnosť zhromažďovať informácie, prípadne presvedčiť príjemcov, aby verili, že existujú neexistujúce udalosti.

Niektoré hoaxy spôsobili falošné poplachy, ktoré spôsobili negatívny dopad na isté organizácie. Jeden z týchto poplachov bol o existencii kyslého dažďa v dôsledku výbuchu jadrového reaktora v Japonsku. Spôsobil chaos v juhovýchodnej Ázii, pretože bol rozosielaný mnohým ľuďom po týždni od zemetrasenia v Japonsku. Ich obsah bol zavádzajúci.

Hoci niektoré hoaxy nie sú škodlivé, často narúšajú chod spoločnosti. Ak je ľudská myseľ infikovaná nepravdivými informáciami, je ťažké človeka presvedčiť o skutočnej pravde.

2 Prístupy na ochranu voči sociálnemu inžinierstvu

Existuje niekoľko prístupov na ochranu voči sociálnemu inžinierstvu. Medzi ne môžeme zaradiť napríklad natural language processing, blacklisty, algoritmy dolovania dát a iné. Vo väčšine existujúcich riešeniach sa využívajú na detekciu práve algoritmy dolovania dát, pretože sa ukázalo, že sú najefektívnejšie.

2.1 Natural Language Processing

Natural Language Processing (NLP) je počítačový prístup k analýze textu, ktorý je založený na teóriách, ale aj na množine technológií. [9] Keďže je to veľmi aktívna oblasť výskumu a vývoja, neexistuje jediná dohodnutá definícia, ktorá by bola vhodná pre každého.

Je to podoblasť počítačovej vedy, informačného inžinierstva a umelej inteligencie zaoberajúcej sa interakciou medzi počítačmi a ľudskými jazykmi, najmä ako programovať počítače na spracovanie a analýzu veľkého množstva údajov jazyka. [9]

Prístupy NLP by sa dali rozdeliť do 4 rôznych kategórií, a to symbolické, štatistické, spojovacie a hybridné prístupy. Symbolické prístupy vykonávajú hĺbkovú analýzu jazykových javov a sú založené na explicitnom zastúpení faktov o jazyku prostredníctvom dobre pochopených schém reprezentácie poznatkov a súvisiacich algoritmov. [9] Štatistické prístupy využívajú rôzne matematické techniky a často používajú veľké textové korpusy na vytvorenie približných zovšeobecnených modelov jazykových javov založených na skutočných príkladoch. [9] Podobne ako štatistické prístupy, aj spojovacie prístupy vyvíjajú zovšeobecnené modely z príkladov jazykových javov.

SEAHound je systém, ktorý vyvinuli autori vo svojej práci [10] za účelom detegovania škodlivých emailov. Tento systém sa zaoberá spracovaním a analýzou textu a je efektívny na detekciu phishingových emailov, ktoré pozostávajú z čistého textu. Dataset, ktorý použili sa skladá z 5009 phishingových emailov a 5000 legitímnych emailov. V emailoch sa zamerali na 4 hlavné atribúty a to či email obsahuje urgentnosť, „zlú“ otázku, škodlivý URL odkaz a milý pozdrav. Na analýzu odkazu použili Netcraft Anti-Phishing Toolbar, čo je komerčný nástroj a ukázal sa byť efektívny už v iných podobných prácach. Presnosť tohto systému je 95%.

V ďalšom článku [11] sa autori zamerali na detekciu phishingových emailov, ktoré neobsahujú odkazy, ale vydávajú sa za banky a snažia sa presvedčiť obeť, aby poskytla citlivé údaje. Navrhnutý systém využíva **NLP** a **WordNet [12]**. WordNet je databáza obsahujúca synonymá rôznych slov. V tejto práci sa zamerali na to či email obsahuje oslovenie menom, či sa spomínajú peniaze, či je email ukončený s prosbou o odpoveď a podobne. Otestovali 600 phishingových emailov a 400 legitímnych emailov, pričom dosiahli úspešnosť 99,4%.

Natural Language Processing bol využitý aj pri vytváraní systému **PhishNet-NLP [13]**. Tento systém operuje medzi užívateľovým mailovým prenosovým agentom a mailovým užívateľovým agentom a spracováva každý email ešte pred tým, než vôbec dorazí do mailovej schránky užívateľa. Extrahovali atribúty (textové, či URL) z emailu pomocou regulárnych výrazov. V databáze s 2000 phishingovými emailami tento klasifikátor správne identifikoval až 98% z nich. Z 1000 legitímnych emailov identifikoval správne až 99,3% z nich.

2.2 K-Nearest Neighbors

Algoritmus **k-NN** je neparametrická klasifikačná technika, ktorá sa ukázala, že je účinná v štatistických aplikáciách rozpoznávania. [14] Tento algoritmus je jednou z najjednoduchších a efektívnych metód pre klasifikáciu. Počíta vzdialenosť medzi testovacou vzorkou a všetkými existujúcimi vzorkami v tréningovej množine. Vytvára rozhodnutie na základe jeho k-najbližších vzorkách po aplikovaní „vážiacej“ funkcie. Vytváranie predikcií s malou množinou údajov má však aj nevýhody. Vzdialenosť medzi každým novým emailom a všetkými existujúcimi emailami by mala byť počítaná ešte pred vytvorením rozhodnutia, čo je výpočtovo zložité pre veľké množstvo emailov.

Saberi a spol. [15] použili súbor klasifikátorov na detekciu scamu. Tento súbor kombinuje k-NN algoritmus, Poissonové rozdelenie a Naive Bayes algoritmy na zlepšenie výkonu systému. Atribúty použité v tejto práci sú všetky textovej povahy. Dataset sa skladá z 529 scamových, 4500 podvodných a 1600 legitímnych emailov. Všetky klasifikátory boli zamerané na text, pričom pozorovali 2400 najčastejšie sa vyskytujúcich termínov, čo spôsobilo neefektívnosť systému, pretože existujú systémy, ktoré používajú na detekciu len štrukturálne atribúty, čo je omnoho rýchlejšie.

Mokhtari, Saraee a Haghshenas [16] sa rozhodli detegovať scam. Tiež navrhli nový prístup na detekciu tým, že vylepšili algoritmus k-Nearest Neigbhors – upravili rovnicu podobnosti dokumentov. Porovnaním klasického algoritmu k-NN a toho vylepšeného, došli k záveru, že nový navrhnutý prístup je o niečo efektívnejší, pretože úspešnosť je 99,05%.

2.3 Support Vector Machines

Princípom **SVM [17]** je rozdelenie tréningových údajov zväčša na dve triedy, no nevylučuje sa ani rozdelenie na viac tried. Cieľom je naučiť sa klasifikačné pravidlo z tréningovej množiny, aby sme mohli v budúcnosti priradiť konkrétnu triedu všetkým novým subjektom.

Pre binárnu klasifikáciu (počet tried je 2) bol tento prístup veľmi dobre vyvinutý. Rozdeľuje oblasť nadrovinou na bodovom diagrame na dve triedy a určuje, ktoré body patria do ktorej triedy. Hlavnou myšlienkou je klasifikovať tak, aby bol okraj (vzdialenosť medzi nadrovinou a najbližším bodom) maximálny. Nás ale zaujíma ako funguje Support Vector Machine pre k > 2, pričom k je počet tried, do ktorých chceme našu údajovú množinu klasifikovať.

Jedna z možností ako riešiť problém, je rozdeliť si ho na množinu binárnych klasifikačných problémov – zostrojí sa k klasifikátorov, jeden pre každú triedu. K-ty klasifikátor skonštruuje nadrovinu medzi poslednou triedou a k-1 inými triedami. Alternatívne môže byť skonštruovaných $\frac{k(k-1)}{2}$ nadrovín, pričom sa oddelí každá trieda od zvyšných. Tieto metódy nazývame „One-against-the-Rest“ a „One-against-One“.

Form a spol. [18] sa problém klasifikácie emailov rozhodli riešiť tak, že vytvorili systém, ktorý pracuje s údajmi z emailu, ktoré sú typu URL, obsahové a tiež sledujú tie atribúty, ktoré môžu meniť správanie počítača. Analyzovali emailovú hlavičku, aby mohli extrahovať informácie ako id správy, email odosielateľa a návratovú cestu. Podľa nich je rozumné analyzovať tieto informácie, pretože môžu byť len ťažko pozmenené útočníkmi. Dokopy z emailu extrahovali 9 konkrétnych atribútov. V tejto práci sa rozhodli kategorizovať emaily len na phishingové alebo legitímne. Dataset poskladali z 2 verejných zdrojov a to z projektu SpamAssasin [19] a Nazario [20] , pričom spolu použili 1000 emailov – 500 legitímnych a 500 phishingových. Na klasifikovanie emailov použili Support Vector Machine, pretože je jeden z najjednoduchších

a najefektívnejších vo viacdimenzionálnych priestoroch a je silný v binárnej klasifikácii. Tento navrhnutý model dosiahol úspešnosť až 97,75%.

Bergholz a spol. [21] navrhli taký systém, v ktorom využívajú kompresiu pomocou dynamického Markovovského reťazca na detekciu phishingového problému a tiež navrhli algoritmus, ktorý redukuje požiadavky na pamäť až o dve tretiny. Tiež navrhli Dirichletov model na túto tému, pričom má lepší výkon ako LDA technika. Z emailov extrahovali podobné údaje ako predošlá práca a teda základné, štrukturálne, URL odkazové a tiež slová, ktoré môžu indikovať, že sa jedná o podvrhnutý email, pričom dokopy vybrali 27 rôznych atribútov z emailu. Na klasifikáciu bol použitý tiež SVM, pričom úspešnosť dosiahli 99,19%.

Modifikáciou klasického SVM na **CS-SVM** (Cuckoo Search Support Vector Machine) **Niu a spol. [22]** získali lepšie výsledky. Extrahovali 23 features rôzneho typu – telové, URL a hlavičkové. Vybudovali hybridný klasifikátor, kde Cuckoo Search je použitý na vyberanie parametrov jadrovej funkcie. Tento kombinovaný klasifikátor prináša omnoho lepšie výsledky ako klasický SVM. Experimentovali na rôznych datasetoch, pričom počty emailov sú 1384 phishingových a 20071 legitímnych emailov. Úspešnosť na tak veľkom datasete bola viac ako 91%.

2.4 Decision Tree

V štruktúre **rozhodovacieho stromu** každý vnútorný uzol označuje test na atribúte, každá vetva predstavuje výsledok testovacích údajov a každý listový uzol identifikuje konkrétnu triedu. Najvyšším uzlom v strome je koreňový uzol. Na vytvorenie rozhodovacieho stromu existujú rôzne algoritmy, ako napríklad CART, ID3, C4.5 atď. Jedná sa o greedy prístup, v ktorom sa rozhodovací strom konštruuje zhora nadol. [23]

ID3 [24] je algoritmus sledovaného učenia, vytvára rozhodovací strom z pevného súboru príkladov. Výsledný strom sa používa na klasifikáciu budúcich vzoriek. Algoritmus ID3 vytvára strom založený na informáciách („Information Gain“) získaných z tréningovej množiny a potom ich používa na klasifikáciu testovacích údajov. Všeobecne používa nominálne atribúty pre klasifikáciu bez chýbajúcich hodnôt.

C4.5 [24] je rozšírenie algoritmu ID3. Tento algoritmus vytvára rozhodovacie stromy veľmi podobne ako ID3. V každom uzle stromu si C4.5 zvolí jeden atribút

údajov, ktorý najúčinnejšie rozdelí svoju množinu vzoriek do podmnožín. Jeho kritériom je normalizovaný zisk informácií, ktorý je výsledkom výberu atribútu na rozdelenie údajov. Na rozhodnutie sa zvolí atribút s najvyššou normalizovanou hodnotou „Information Gain“.

Algoritmus **C5.0 [24]** je opäť len rozšírenie algoritmu C4.5, pričom je rýchlejší, môže využívať počítače s viacerými procesormi alebo jadrami, zlepšuje presnosť predikcie a podobne.

Toolan a Carthy sa vo svojom výskume [25] zaoberali selekciou atribútov pre detekciu spamu a phishingu. Z každého emailu extrahovali 40 atribútov a tie zahŕňali informácie o tele emailu, predmete emailu, URL odkazoch, odosielateľovi a či email obsahuje spúšťacie skripty. Vytvorili 3 rôzne datasety, pričom prvý obsahoval legitímne emaily a spam, druhý obsahoval legitímne emaily a phishingové emaily a tretí obsahoval všetky vyššie spomenuté emaily. Taktiež ohodnotili dôležitosť každého atribútu pomocou „Information Gain“ a ako klasifikátor použili C5.0 rozhodovací strom, pričom výsledkom bolo, že tento rozhodovací strom poskytol najlepšie výsledky, ak sa pracovalo s atribútmi, ktoré mali najvyššiu hodnotu „Information Gain“.

2.5 Random Forest

Náhodný les je výpočtovo efektívna technika, ktorá dokáže rýchlo pracovať s veľkými súbormi údajov. [26] Prináša veľmi vysokú presnosť klasifikácie, nový spôsob určenia variabilného významu a schopnosť modelovať komplexné interakcie medzi prediktorovými premennými. Je veľmi flexibilný pri vykonávaní niekoľkých typov štatistických analýz údajov vrátane regresie, klasifikácie a podobne.

Náhodné lesy sa skladajú z kolekcie klasifikátorov stromovej štruktúry, pričom každý strom závisí od hodnôt náhodného vektora. Generalizačná chyba lesa klasifikačných stromov závisí od sily jednotlivých stromov v lese a korelácie medzi nimi. Sú rozšírením Breimanovho [27] bagging nápadu a boli vyvinuté kvôli zlepšeniu výkonu. Sú lákavé práve z dôvodu, že sú flexibilné, sú relatívne rýchle na trénovanie a následné predikovanie, závisia od málo ladiacich parametrov, môžu byť použité na viacdimenzionálne problémy a sú veľmi ľahko implementovateľné.

Yasin a Abuhasan [28] vybudovali inteligentný klasifikátor, ktorý je schopný detegovať phishingové emaily. Využívajú techniky dolovania dát, pričom ich model sa

učí z tréningového datasetu, ktorý obsahuje legitímne aj phishingové emaily. Model extrahuje 16 atribútov z emailov, ktorým je pridelená hodnota dôležitosti. Na tomto modeli otestovali a následne porovnali 5 známych klasifikačných techník zahŕňajúc J48, Naive Bayes, Support Vector Machine, Multilayer Perceptron a Random Forest. Najlepšie výsledky boli dosiahnuté pomocou Random Forest, aj keď rozdiely neboli príliš veľké. Dataset obsahoval 5940 legitímnych emailov a 4598 phishingových emailov.

Ďalší systém, ktorý bol vytvorený, bol nazvaný **PFILFER [29]**. Klasifikuje phishingové emaily a tak ako v predošlých podobných prácach, aj v tejto autori rozdeľujú emaily len do dvoch tried – či sa jedná o phishing alebo nie. Extrahovali 10 atribútov, ktoré zväčša prihliadali na URL odkazy, poprípade či email obsahuje Javascript. Použili približne 7000 legitímnych a 860 phishingových emailov. V najlepšom prípade dosiahli úspešnosť 99,5% pomocou algoritmu Random Forest.

Abu-Nimeh a spol. [30] sa rozhodli porovnávať 6 najčastejšie používaných klasifikátorov na phishingovom datasete. Klasifikátory zahŕňajú logistickú regresiu, CART, SVM, RF a neurónové siete. Dataset obsahuje 1171 phishingových emailov a 1718 legitímnych emailov. Z emailov extrahovali 43 hodnôt atribútov, pričom pri použití 10-fold krížovej validácie Random Forest predbehol všetky ostatné a úspešnosť bola 92,28%.

2.6 Iné

Hamid a spol. [31] priniesli model, ktorý na základe 8 hybridných atribútov vie identifikovať phishingový email. Zamerali sa na tie údaje, ktoré útočník nevie pozmeniť. Email, ktorý zvykne prichádzať z viacerých domén vyvoláva podozrenie, že sa jedná o abnormálnu aktivitu. Porovnali viacero algoritmov (Bayesova sieť, SVM, Adaboost, Random Tree), pričom najlepšie výsledky dosiahli pomocou algoritmu Bayesovej siete.

Sharma a spol. [23] sa rozhodli, že budú porovnávať 2 klasifikátory na detekciu spamu - Naive Bayes a Multi Layer Perceptron (MLP). Podľa ich výsledkov je MLP lepší klasifikátor (úspešnosť je 93%), ale na druhej strane, stavba modelu pre MLP zabrala až 78krát viac času.

3 Prípravná fáza pre zatrieďovanie emailov

V tejto časti práce sa venujeme prípravnej fáze pre následné zatrieďovanie emailov do konkrétnych kategórií. V prvom rade bolo potrebné pripraviť dataset, s ktorým by sme mohli pracovať. Porovnali sme najčastejšie používané algoritmy v podobných prácach.

3.1 Príprava údajov

Na vytvorenie efektívneho systému sme potrebovali dataset, ktorý by mal kategorizované emaily. Taký dataset ale vo všeobecnosti nie je, a preto sme z projektu **Enron-Spam [32]** stiahli dataset, z ktorého sme použili 298 emailov, pričom sme ich museli kategorizovať. 150 emailov je legitímnych a zvyšných 148 sú podvrhnuté. Niekoľko z týchto podvrhnutých emailov sme zozbierali aj od ľudí z laboratória kybernetickej bezpečnosti na UPJŠ.

Zo začiatku sme pracovali s datasetom, ktorý bol rozdelený mnou do viacerých kategórií. Problémom bolo, že mohla nastať veľká chybovosť kvôli rôznym okolitým faktorom (ruch, únava a podobne) a teda sme sa rozhodli, že by bolo vhodnejšie, keby tieto emaily roztriedili viacerí ľudia.

Vytvorili sme formulár, ktorý obsahoval niekoľko vybraných podvrhnutých emailov z datasetu. Požiadali sme ľudí z akademického CSIRT tímu, ktorých sme vopred naučili ako majú odlíšiť rôzne kategórie sociálneho inžinierstva, aby nám pomohli roztriediť tieto emaily. 63 emailov bolo zaradených do kategórie spam, 25 do kategórie phishing, 4 boli spearphishingové a 56 padlo do kategórie scam. Ani jeden z emailov nebol označený ako hoax. Práve preto sme sa rozhodli zrušiť kategóriu hoax a tiež spearphishing, pričom 4 získané spearphishingové emaily sme pridali do kategórie phishing. Tieto výsledky sa len veľmi málo odlišujú od toho, ako to bolo zatriedené predtým. Minimálne sme sa uistili, že sme emaily roztriedili korektne.

Náš dataset sa bude postupne napĺňať emailami, ktoré sú cielené priamo na našu univerzitu a teda sa celkovo aj systém prispôsobí na prácu s univerzitnými emailami. Zatiaľ sme pracovali s údajmi, ktoré sme mali dostupné z internetu. Na základe týchto kategorizovaných emailov budeme môcť neskôr klasifikovať ďalšie emaily, na ktoré sa používatelia budú dopytovať. Dataset, s ktorým sme pracovali je súčasťou prílohy A.

3.2 Výkonnostná metrika

Aby sme zistili, ako presný model sme vytvorili, určili sme si výkonnostnú metriku, na základe ktorej vieme percentuálne vyjadriť úspešnosť a iné hodnoty nášho modelu pre každý algoritmus zvlášť.

Pre jednoduchšiu orientáciu nazvime všetky podvrhnuté emaily pozitívna trieda a legitímne emaily negatívna trieda. Túto metriku budeme používať len na vyjadrenie hodnôt rozdelenia na legitímne emaily a všetky ostatné.

True Positive (TP) je počet záznamov, kde model správne predpovedá pozitívnu triedu.

True Negative (TN) je počet záznamov, kde model správne predpovedá negatívnu triedu.

False Positive (FP) je počet záznamov, kde model nesprávne predpovedá pozitívnu triedu.

False Negative (FN) je počet záznamov, kde model nesprávne predpovedá negatívnu triedu.

Presnosť (Precision) meria percentuálne vyjadrenie toho, v koľkých prípadoch klasifikátor správne identifikoval pozitívnu triedu ku všetkým identifikovaným prvkom pozitívnej triedy. $Precision = \frac{TP}{TP+FP}$

Citlivosť (Recall) meria úplnosť výsledkov klasifikátora, teda koľko prvkov pozitívnej triedy, bolo naozaj označených ako pozitívna trieda a určujeme to takto: $Recall = \frac{TP}{TP+FN}$

F-skóre je definované ako harmonický priemer presnosti a citlivosti: $F - skóre = \frac{2*Precision*\text{Recall}}{Precision+\text{Reca}}$, ale to pre nás nie je dôležité, pretože nevyjadruje relevantnú hodnotu, ktorá by hovorila napríklad o počte správne klasifikovaných podvrhnutých emailoch a podobne.

Úspešnosť (Accuracy) je percentuálne vyjadrenie pomeru počtu korektne detegovaných prvkov pozitívnej triedy ku všetkým prvkom, ktoré boli na vstupe. $Accuracy = \frac{TP+TN}{TP+TN+FP+F}$

3.3 Porovnanie algoritmov dolovania dát

Na dosiahnutie čo najlepších možných výsledkov sme sa rozhodli porovnať niekoľko často používaných algoritmov na klasifikáciu, aby sme v systéme využili ten s najlepšími výsledkami.

V prvom rade je dôležité či náš systém správne identifikuje legitímny a podvrhnutý email, až potom tieto emaily klasifikujeme do konkrétnych tried. Najprv si teda ukážeme, či sme vhodne zvolili atribúty na identifikáciu legitímneho a podvrhnutého emailu. Na ohodnotenie použijeme vyššie spomenutú výkonnostnú metriku, pričom porovnáme všetky hodnoty pre každý algoritmus.

Nad datasetom sme porovnali 4 algoritmy a to **Random Forest, Support Vector Machines, k-Nearest Neighbors** a **Decision Tree**, ktorých implementáciu nám ponúkla knižnica scikit-learn [33] v programovacom jazyku Python 3.7 [34]. Po niekoľkých testoch sme si určili, že tréningová množina údajov bude 60% celého datasetu a testovacia množina údajov bude 40% datasetu, teda sme testovali ako algoritmy roztriedia 120 emailov. Veľmi dôležitá hodnota je koľko podvrhnutých emailov klasifikátor určil, že sú naozaj podvrhnuté (TP) a v koľkých sa pomýlil a identifikoval ich ako legitímne.

Tieto algoritmy roztriedili emaily na základe atribútov veľkosť emailu, počet príloh, počet URL, maximálny počet bodiek, počet URL v tvare IP adresy, počet URL v tvare obrázku, počet URL so @, počet URL s %, počet URL s neštandardným portom, maximálna dĺžka URL, počet URL, ktoré sú rôzne pre tag „a href“ a text za ním a tiež prítomnosť najčastejšie vyskytujúcich sa slov. Viac si ich priblížime v kapitole 4.3 Extrakcia atribútov z emailu.

Vo všeobecnosti pre nás nie je dôležité, pokiaľ algoritmus má najlepšie výsledky v jednom prípade. Po použití **10-fold krížovej validácie** sme vyhodnotili, že Random Forest by mohla byť najlepšia možná voľba. V tab. 1 si môžeme pozrieť priemernú úspešnosť jednotlivých algoritmov.

Tab. 1 Porovnanie priemernej úspešnosti algoritmov dolovania dát

	RF	**kNN**	**DT**	**SVM**
Úspešnosť	0.8940518575	0.7690247678	0.7685758513	0.8566013071

Pre jedno konkrétne meranie Random Forest vyhodnotil, že TP je až 61 zo 65 emailov, pričom pri kNN to bolo 52 zo 65, pri DT bolo TP 59 zo 65 a pre SVM sme dostali hodnotu 52 z 57. V tab. 2 vidíme výsledky pre jedno konkrétne meranie:

Tab. 2 Porovnanie algoritmov dolovania dát

	RF	**kNN**	**DT**	**SVM**
TP	61 zo 65	52 zo 65	59 zo 65	52 z 57
TN	49 z 55	55 z 55	48 z 55	55 zo 63
FP	6 z 55	0 z 55	7 z 55	8 zo 63
FN	4 zo 65	13 zo 65	6 zo 65	5 z 57
Presnosť	0,9104478	1	0,8939394	0,8666667
Citlivosť	0,9384615	0,8	0,9076923	0,9122807
Úspešnosť	0,9166667	0,8916667	0,8916667	0,8916667

Jednotlivé algoritmy – RF, k-NN, DT a SVM v tomto konkrétnom prípade rozdelili emaily do nami určených kategórií, pričom v stĺpcoch je očakávaná trieda a v riadku predikovaná. Výsledky validácie môžeme vidieť v tab. 3, tab. 4, tab. 5 a tab. 6:

Tab. 3 Vyhodnotenie algoritmu Random Forest

RF		**Predikované triedy**			
		legitímny	**phishing**	**scam**	**spam**
Očakávané triedy	**legitímny**	49	2	2	2
	phishing	1	4	1	7
	scam	2	0	20	2
	spam	1	1	0	26

Tab. 4 Vyhodnotenie algoritmu k-Nearest Neighbors

k-NN		**Predikované triedy**			
		legitímny	**phishing**	**scam**	**spam**
Očakávané triedy	**legitímny**	55	0	0	0
	phishing	3	1	0	9

	scam	5	0	18	1
	spam	5	0	0	23

Tab. 5 Vyhodnotenie algoritmu Decision Tree

DT		**Predikované triedy**			
		legitímny	**phishing**	**scam**	**spam**
Očakávané triedy	**legitímny**	48	1	2	4
	phishing	1	5	1	6
	scam	4	1	18	1
	spam	1	2	0	25

Tab. 6 Vyhodnotenie algoritmu Support Vector Machines

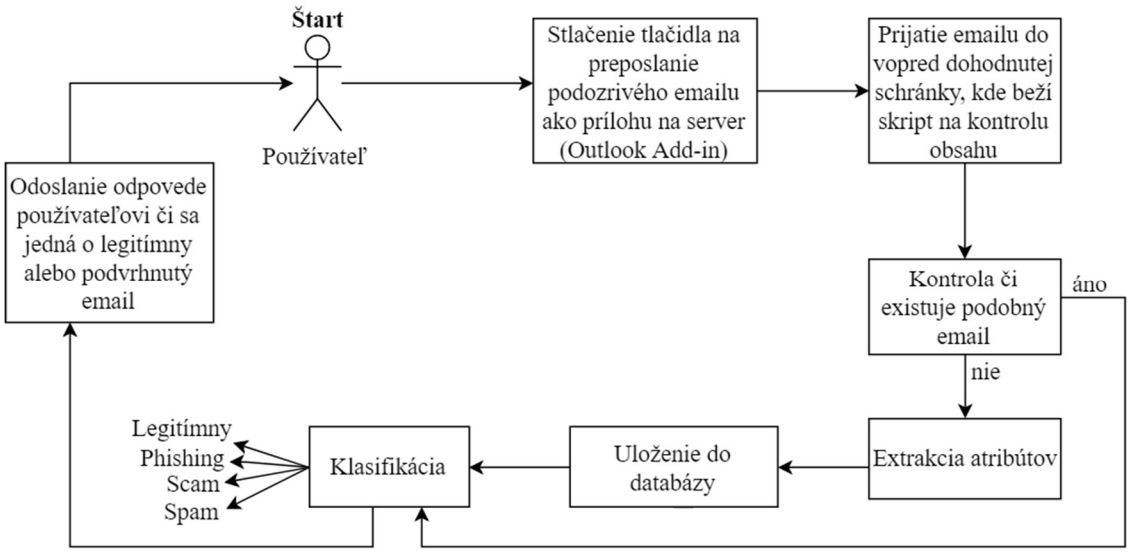
SVM		**Predikované triedy**			
		legitímne	**phishing**	**scam**	**spam**
Očakávané triedy	**legitímne**	55	3	2	3
	phishing	2	5	2	1
	scam	2	0	23	0
	spam	1	2	1	18

Môžeme si všimnúť, že každý z algoritmov má problém korektne klasifikovať do kategórie phishing, pri tejto triede dochádza k najväčšej chybovosti. Zjavne sa nedá úplne presne určiť či sa jedná o phishing. Dôvodom môže byť aj to, že na vstupe mali tieto algoritmy málo phishingových emailov, z ktorých by sa mohli učiť. Pre nás je dôležité či ten email je klasifikovaný ako podvrhnutý, a teda padne do kategórie phishing, scam alebo spam.

Ukázalo sa teda, že pokiaľ chceme, aby náš systém vyhodnocoval nové emaily čo najlepšie, tak by bolo vhodné v ňom použiť práve **Random Forest**. Za zmienku stojí aj SVM, poprípade je tu možnosť kombinácie všetkých algoritmov. Teda by sme sa pozreli, ako rozhodol každý jeden a na základe toho vyhodnotiť konečnú triedu, do ktorej by email padol.

4 Systém na detekciu foriem sociálneho inžinierstva v emailovej komunikácii

Celý systém sa skladá z viacerých častí, a to z modulu na preposielanie podozrivého emailu na server, monitoringu prichádzajúcich emailov v aplikácii – v tom je zahrnutá kontrola existencie podobného emailu, z extrakcie atribútov z emailu v prípade, že sme nenašli v databáze podobný email, z databázy, do ktorej ukladáme extrahované atribúty a požiadavky, z klasifikácie emailu, administrátorského rozhrania a nakoniec z modulu na odosielanie spätnej väzby používateľovi. Manuál na používanie tohto systému je v prílohe B a v elektronickej forme tiež v prílohe A. Schéma systému je na obr. 2:



Obr. 2 Schéma systému na detekciu foriem sociálneho inžinierstva

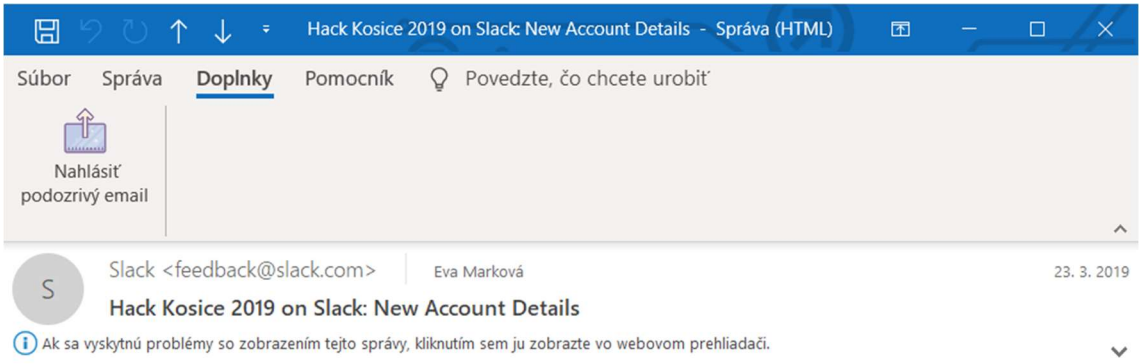
4.1 Modul na preposielanie podozrivého emailu na server

V prípade, že používateľ obdrží email, pričom si nie je istý, či náhodou tento email nie je podvrhnutý, mal by mať možnosť preposlať tento email na vopred dohodnutú adresu. Na základe tejto požiadavky sme sa rozhodli implementovať doplnok do emailového klienta, pričom stačí, aby používateľ klikol na tlačidlo, o ostatné sa už postará kód v pozadí.

Najlepšou možnou voľbou je **Add-in do Outlooku [35]**, keďže v rámci UPJŠ zamestnanci využívajú konto, ktoré im poskytla univerzita. Súčasťou tohto konta je aj

balík Office365 [36], pričom tento balík obsahuje aplikáciu Outlook a teda zamestnanci používajú práve toto rozhranie na emailovú komunikáciu vo väčšine prípadov.

Tento Add-in do Outlooku bol implementovaný v jazyku C#, pričom Visual Studio 2017 [37] ponúka priamo vytvorenie projektu typu „Office/SharePoint“ a následne typu „Outlook 2013 and 2016 VSTO Add-in“. V tejto šablóne sa priamo vytvorí trieda, ktorá má na starosť, čo sa má udiať hneď po spustení/vypnutí aplikácie. V našom prípade sme vytvorili tlačidlo, ktoré po stlačení vezme obsah emailu, vrátane hlavičky a príloh a prepošle ho ako prílohu s automaticky vygenerovaným predmetom a obsahom emailu. Na obr. 3 môžeme vidieť ukážku Add-inu do Outlooku.



Obr. 3 Ukážka Add-inu do Outlooku

4.2 Monitoring prichádzajúcich emailov v aplikácii

V rámci aplikácie Outlook sme vytvorili skript, ktorý kontroluje či nepribudol nový email na aktuálne prihlásenom účte. Pokiaľ email, ktorý bol identifikovaný ako nový, má požadované znenie predmetu, otvorí sa príloha – teda email, o ktorom si používateľ nie je istý či je legitímny, alebo podvrhnutý a následne prebehne kontrola existencie podobného emailu.

4.2.1 Kontrola existencie podobného emailu

Na kontrolu či existuje podobný email sme použili v Pythone 3.7 knižnicu „**fuzzywuzzy**“ [38], ktorá nám dala k dispozícii funkciu „ratio“. Tá nám vracia hodnotu v percentách, na koľko sú si dva emaily podobné. V implementácii využíva Levenstheinovu vzdialenosť [39] na zistenie podobnosti. Aj keď sa zdá, že by táto metóda mohla mať isté nevýhody, hlavne výkonnostné, tak sa nám v praxi ukázalo, že

funguje dostačujúco a spoľahlivo. Predstavme si prípad, že v databáze máme uložených už niekoľko stoviek údajov a podobný email je niekde medzi poslednými. To znamená, že hľadanie podobného emailu predstavuje prejdenie niekoľkých stoviek riadkov, pričom v každom z nich musí počítať vždy novú hodnotu.

Lepšou voľbou je „**ssDeep**“ [40], pričom sa jedná o vytváranie nie celkom klasického hash-u. V prípade zmeny nejakej časti textu je nový hash stále podobný tomu predošlému, a teda tieto dva hashe sú porovnateľné. V podstate stačí raz vytvoriť hash každého emailu a uložiť ho do databázy. Potom už porovnávame len hashe týchto emailov, čo je omnoho rýchlejšie ako porovnávať podobnosť dvoch textov. Tento „ssDeep“ sa pôvodne používal na porovnávanie súborov na identifikáciu vírusov, a preto nie je celkom triviálne používať ho na porovnávanie dvoch emailov. Preto sme sa zatiaľ rozhodli použiť vyššie spomínanú knižnicu „fuzzywuzzy“ a v budúcej práci sa možno zameriame práve na tento „ssDeep“.

Po testovaní ako sa zmení hodnota podobnosti pri nejakých zmenách emailov, sme si stanovili **prah 70%**. Pokiaľ je miera podobnosti nižšia, nemôžeme si byť istí, že sa stále jedná o podobný email. Radšej takýto email klasifikujeme nanovo, aby naše výsledky boli presnejšie. Ak je hodnota podobnosti vyššia alebo rovná tomuto prahu, vieme s istotou povedať, že sa v databáze vyskytuje podobný email a teda podľa neho vieme určiť kategóriu emailu, ktorý sme chceli klasifikovať.

4.3 Extrakcia atribútov z emailu

V prípade, že nie je možné na základe podobnosti povedať, o akú kategóriu sa jedná, je potrebné z emailu extrahovať atribúty, vďaka ktorým budeme vedieť email klasifikovať. Po naštudovaní problematiky sme sa rozhodli vynechať hlavičkové atribúty. Dnes už nie je dôležité spracovávať napríklad čas odoslania, pretože emaily sociálneho inžinierstva sa odosielajú v rozmanitých časoch.

Na spracovávanie a kategorizovanie emailov sú dôležité základné informácie o emailoch, atribúty extrahované z URL odkazov a tiež najčastejšie vyskytujúce sa slová, pretože napríklad scam nezvykne obsahovať URL odkazy a pomocou základných informácií by sme len ťažko identifikovali, že sa jedná o scam.

Dokopy extrahujeme **11 atribútov** a zároveň prihliadame na frekvenciu výskytu **35 konkrétnych slov**, čo je pre nás kľúčové pri klasifikácii do konkrétnej kategórie.

4.3.1 Základné informácie o emaile

Je potrebné sa zamerať na základné informácie o emaile, a teda sme sa rozhodli extrahovať dve dôležité informácie, ktoré by mohli pomôcť pri identifikovaní podvrhnutého emailu a to veľkosť emailu a počet príloh:

- **Veľkosť emailu** – veľmi dôležitý parameter vzhľadom k tomu, že nám môže napovedať, či email obsahuje prílohu a podobne.
- **Počet príloh** – prítomnosť prílohy môže v kombinácii s inými parametrami indikovať, že email je nebezpečný.

4.3.2 Atribúty extrahované z URL

V emailoch sociálneho inžinierstva sa často vyskytujú škodlivé odkazy a teda to môžu byť dobrým indikátorom, že sa jedná o podvrhnutý email. Z tohto hľadiska sme sa pozreli na niektoré atribúty súvisiace s URL odkazmi:

- **Počet URL** – škodlivé emaily často obsahujú niekoľko URL odkazov na podvrhnuté webové stránky, a preto je pre nás počet URL dôležitý atribút.
- **Maximálny počet bodiek** – veľa útočníkov sa snaží skonštruovať legitímne vyzerajúci URL odkaz tak, že pridá niekoľko už známych domén, napríklad http://secure.login.paypal.com/ môže v užívateľovi evokovať, že sa jedná o legitímnu stránku.
- **Počet URL v tvare IP adresy** – niektoré škodlivé stránky môžu byť hostené na počítačoch, ktoré nemusia mať DNS záznamy. DNS prekladá doménové meno, ktoré napíšeme do prehliadača na IP adresu webového servera, ktorý je hostiteľom webovej stránky. A teda najjednoduchšia cesta ako detegovať škodlivé stránky je pomocou IP adresy.
- **Počet URL v tvare obrázku** – jedná sa o obrázky vyskytujúce sa v tagu „a href“.
- **Počet URL so @** – pokiaľ URL odkaz obsahuje @, webový prehliadač odignoruje tú časť pred týmto symbolom a načíta sa adresa, ktorá je za týmto znakom.

- **Počet URL s %** – útočník môže zakódovať časť URL odkazu hexadecimálnymi znakmi za účelom ukrytia podvodnej webovej adresy. Takáto adresa môže vyzerať napríklad takto: http://fast-visa.com\%32\%34\%2c, pričom prehliadač zobrazí len legitímnu časť v paneli.
- **Počet URL s neštandardným portom** – zisťujeme počet URL s portom rôznym od 80 (http) a 443 (https). Každý iný port je pre nás podozrivý.
- **Maximálna dĺžka URL** – útočníci často používajú dlhé URL odkazy, aby skryli tú pochybnú časť.
- **Počet URL, ktoré sú rôzne pre tag „a href" a text za ním** – ak je rozdiel medzi „a href" atribútom a textom za ním, tak nadobúdame podozrenie, že môže ísť o škodlivý odkaz.

4.3.3 Najčastejšie vyskytujúce sa slová

Vzhľadom k tomu, že v našej práci chceme podvrhnuté emaily kategorizovať, je potrebné sa pozrieť aj na frekvenciu výskytu slov. Hlavný dôvod je ten, že v podvrhnutých emailoch sa nemusia vyskytovať len URL odkazy, a teda by sme nevedeli korektne klasifikovať napríklad scam.

Slová, ktoré sme sa rozhodli extrahovať, nie sú náhodné. Nad každou kategóriou sociálneho inžinierstva v našom pripravenom datasete sme spustili vyhľadávanie najčastejšie sa vyskytujúcich slov, pričom počet výskytov má byť viac ako 10 a dĺžka slova aspoň 5, keďže chceme odstrániť slová typu „the","and" a podobne. Keďže scamové emaily obsahujú siahodlhé texty, tam sme sa rozhodli parameter pre počet výskytov zvýšiť až na 25, aby sme zbytočne nesledovali príliš veľa slov. Tab. 7 zobrazuje, koľkokrát sa dané slovo vyskytlo v datasete podľa kategórií.

Tab. 7 Frekvencia výskytu slov v jednotlivých kategóriách

Slovo	Legitímny	Phishing	Spam	Scam
customer	12	4	17	4
viagra	0	0	15	0
information	75	14	21	40
million	7	1	14	67
cialis	3	0	24	0

financial	11	2	11	10
company	6	0	17	81
security	2	6	0	66
statement	4	0	16	32
report	90	4	13	14
investment	2	3	12	28
price	13	3	36	12
account	33	38	3	110
access	19	15	6	6
please	186	23	9	76
change	47	11	6	10
program	49	8	2	18
request	31	7	0	20
review	51	5	4	2
address	15	7	1	55
number	39	4	5	67
online	5	12	4	2
paypal	0	12	0	0
message	43	10	26	8
business	55	6	6	77
assistance	6	1	1	55
foreign	0	0	0	86
transaction	21	0	1	96
country	0	1	0	86
contact	28	12	0	70
nigeria	0	0	0	80
transfer	7	4	1	63
family	3	2	2	59
government	1	0	0	52
contract	19	0	1	75

V prácach [17][20][27] a podobne sa rozhodli extrahovať aj tieto ďalšie atribúty, napríklad hlavičkové, my sme ich však nepovažovali za relevantné, pretože v našej testovacej vzorke sa vyskytovalo málo emailov, z ktorých by tieto údaje boli čitateľné:

- **Doména odosielateľa** – meria podobnosť medzi doménou odosielateľa a doménou message-ID.
- **Unikátny odosielateľ** – reprezentuje či odosielateľ posiela email z viac ako jednej domény.

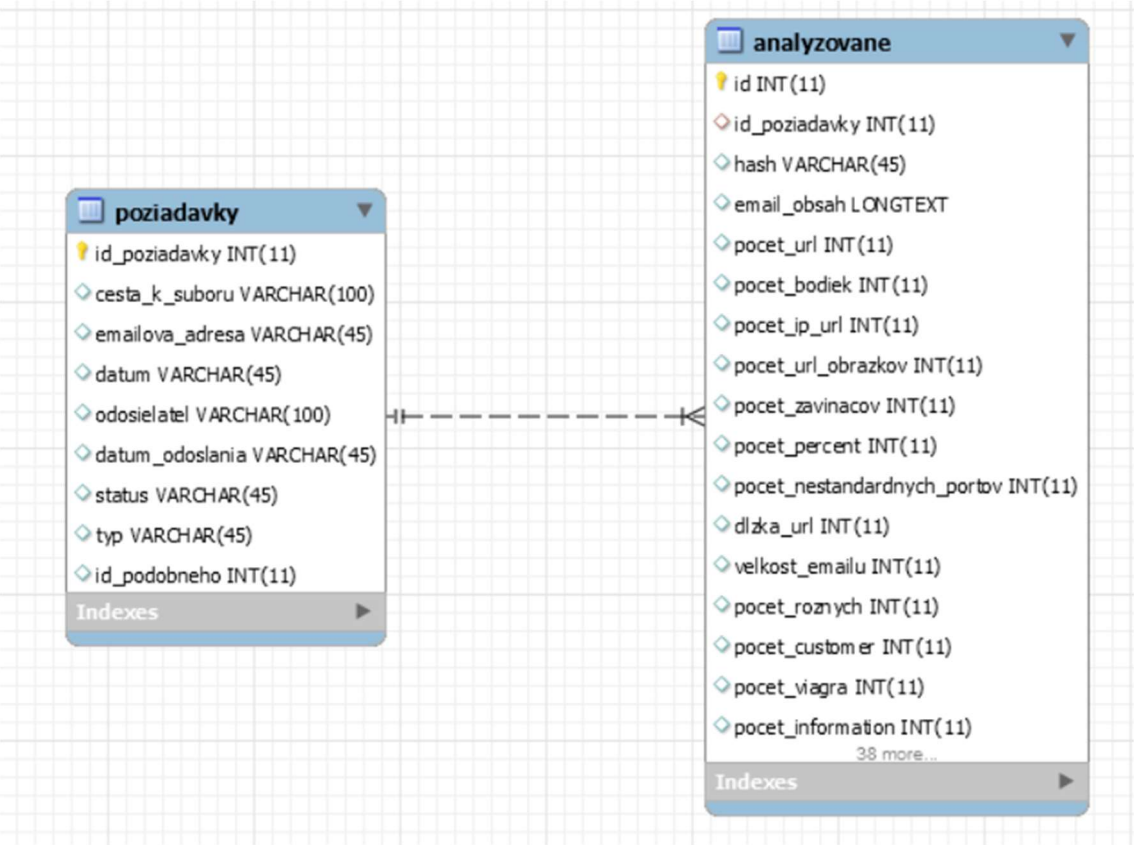
- **Unikátna doména** – kontroluje, či doménové meno je použité viacerými odosielateľmi.
- **Spätná cesta** – vracia 0 alebo 1, podľa toho, či doménové meno spätnej cesty je to isté ako doménové meno odosielateľa.
- **Čas odoslania** – legitímne emaily sú častokrát odosielané legitímnymi odosielateľmi v pracovnom čase, zatiaľ čo útočníci útočia mimo pracovného času, aby predišli pozornosti detektorov.
- **Použitie skracovacích služieb** – zisťuje, či sa v URL odkazoch vyskytujú najznámejšie skracovacie služby ako napríklad https://goo.gl/, https://bitly.com/ a podobne.
- ...

4.4 Databáza

V rámci nášho systému sme sa rozhodli použiť databázu MySQL [40], pričom obsahuje 2 tabuľky. V jednej sa zameriavame na evidenciu požiadaviek od používateľov, pričom do nej ukladáme tieto údaje: cestu k emailu, dátum prijatia požiadavky, emailovú adresu nahlasovateľa, dátum odoslania emailu, email odosielateľa, stav požiadavky, vyhodnotený typ a poprípade id podobného emailu.

V druhej tabuľke si ukladáme analyzované požiadavky, pričom medzi stĺpce, ktoré vyhodnocujeme patria id požiadavky, aby sme vedeli prepojiť tieto dve tabuľky, telo emailu, atribúty, ktoré sme extrahovali z emailov, typ, ktorý sa vyhodnotí pomocou algoritmu RF a tiež hodnoty získané pomocou algoritmov SVM, k-NN a DT. V tom lepšom prípade by sme v tejto tabuľke chceli mať aj hash emailu pomocou „ssDeep-u“, zatiaľ ale pracujeme len s podobnosťou celého obsahu emailu.

Na obr. 4 je schéma databázy:



Obr. 4 Schéma tabuliek v databáze

Na obrázku môžeme vidieť, že tieto dve tabuľky sú prepojené, pričom medzi nimi existuje cudzí kľúč, ktorý sme v oboch tabuľkách pomenovali „id_poziadavky".

4.5 Klasifikácia emailu

Na základe porovnania algoritmov sme sa rozhodli, že v našom systéme využijeme na klasifikovanie nových záznamov algoritmus **Random Forest**, pretože vo viacerých prácach nadobúdal najlepšie výsledky. Určili sme si 4 rôzne kategórie, do ktorých tento algoritmus môže zaradiť nový email. Medzi ne patria spam, scam, phishing a legitímne emaily.

4.6 Administrátorské rozhranie

Na to, aby sme nezanedbali **ľudský faktor**, je potrebné sprístupniť celú databázu administrátorovi, ktorý by mal právo meniť rozhodnutia klasifikátora. Môže sa totižto stať, že klasifikátor nesprávne rozhodne a administrátor si to všimne, teda to bude môcť

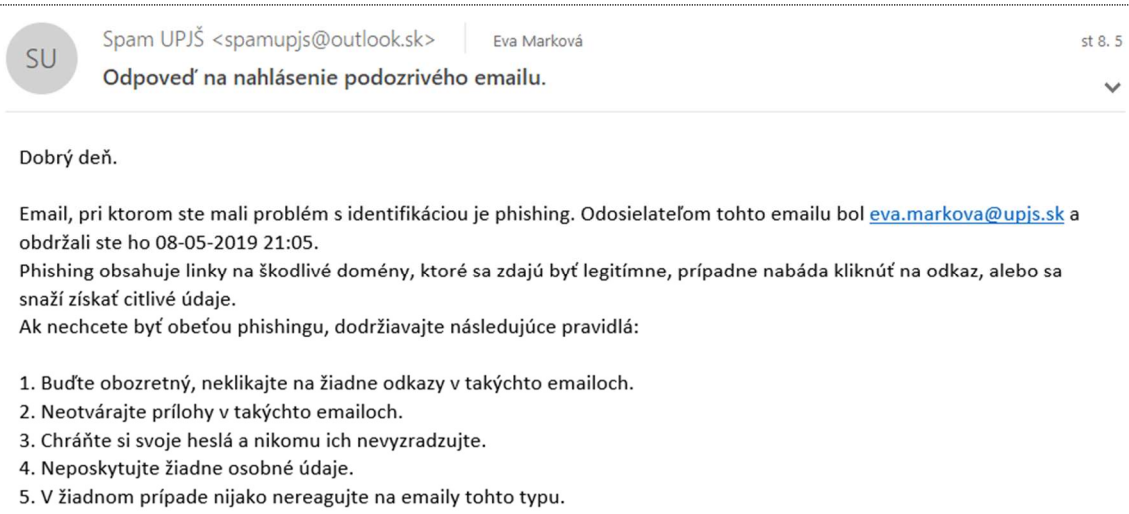
prepísať. Prioritou celého systému je ho zautomatizovať, ale v krajných prípadoch je potrebný aj zásah užívateľa, pretože technológia stále nie je stopercentná a tiež nás môže prekvapiť email, ktorý nebude na 100% klasifikovateľný.

Nad celým systémom má dohľad administrátor, ktorý môže meniť rozhodnutia klasifikátora. V grafickom rozhraní má administrátor k dispozícii výsledky zo všetkých štyroch vyššie spomínaných algoritmov – RF, kNN, SVM a DT (CART). Taktiež vidí náhľad emailu a jeho úlohou je rozhodnúť sa, či súhlasí s výsledkom algoritmu random forest, alebo radšej zvolí inú kategóriu. V tomto rozhraní má administrátor k dispozícii celú históriu požiadaviek (vyriešených aj nevyriešených).

Týmto krokom sme nezanedbali ľudský faktor. Systém je plne automatizovaný v prípade, že už existuje v databáze podobný email. V prípade, že neexistuje, je potrebný zásah administrátora. Následne po kategorizácii administrátorom sú všetky podobné emaily klasifikované automaticky.

4.7 Modul odosielania spätnej väzby

Posledným modulom v rámci tohto systému je odosielanie spätnej väzby používateľovi, ktorý poskytol email, ktorým si nie je istý. Práve kvôli tomuto modulu sme si ukladali do databázy aj emailovú adresu používateľa. Užívateľ dostane odpoveď na svoju otázku a teda sa dozvie, či ide o legitímny alebo podvrhnutý email a kroky, ktoré by mal následne vykonať vzhľadom k identifikovanej kategórii – a teda či email odignorovať, poprípade vymazať alebo s čistým svedomím naňho odpísať. Pre každú kategóriu sme vytvorili špecifickú odpoveď, aby užívateľ vedel, ako sa zachovať v prípade, že sa jedná o scam, spam, phishing alebo legitímny email. Odpoveď, v prípade, že sa jedná o phishing môžeme vidieť na obr. 5. Odpovede sme tvorili na základe získaných skúseností z riešenia týchto udalostí v Centre aplikovanej informatiky na univerzite.



Obr. 5 Odpoveď užívateľovi v prípade, že sa jedná o phishing

V odpovedi, pokiaľ sa jedná o phishing sa snažíme užívateľovi pripomenúť to, čím by sa mal riadiť, teda neklikať na odkazy v emailoch, neotvárať prílohy, nereagovať na emaily podobného typu a podobne. Pokiaľ užívateľ nahlásil scam, odporúča sa mu neposkytovať žiadne bankové údaje a ani nič neplatiť. V prípade, že ide o spam zdôrazňujeme, že môže ísť o vyžiadanú či nevyžiadanú reklamu, a že sa neodporúča klikať na odkazy, pretože môžu byť škodlivé. Pokiaľ užívateľ nahlásil legitímny email, v odpovedi je napísané, že môže bez strachu reagovať na takúto správu.

5 Vyhodnotenie

Na to, aby sme mohli vyhodnotiť systém ako celok sme potrebovali ďalších niekoľko emailov, aby sme otestovali nakoľko je úspešný.

Najprv sme sa rozhodli do systému vložiť **15 podobných emailov**, pričom sme nemenili základnú štruktúru emailu, len zopár slov. Ukázalo sa, že systém rozpoznal, že tieto emaily sú podobné a teda automaticky odoslal odpoveď nahlasovateľovi, o akú triedu sa jedná, a ako sa má užívateľ zachovať.

V ďalšom teste sme si zvolili približne 10% celkového počtu emailov v datasete, konkrétne **29 emailov**, pričom 7 bolo legitímnych a zvyšné podvrhnuté. Jedná sa o úplne nové emaily, ktoré sa v našom datasete ešte nevyskytli. Náhodne sme vybrali emaily, dali sme ich na vstup systému a sledovali sme, ako systém tieto emaily zatriedi. Ukázalo sa, že keďže dataset nie je vyvážený, SVM a k-NN neklasifikujú emaily korektne a prikláňajú sa k väčšinovej triede, preto sme sa rozhodli sledovať výsledky algoritmu **RF** a **DT (CART)**. Pokiaľ jeden z týchto algoritmov rozhodol správne, považujeme email za správne klasifikovaný. Podstatné je, či algoritmus korektne klasifikoval email za legitímny alebo podvrhnutý. V tab. 8 môžeme vidieť výsledky porovnania:

Tab. 8 Porovnanie algoritmov dolovania dát pri testovaní systému

	RF	**DT (CART)**
TP	19 zo 22	16 zo 22
TN	1 z 7	7 z 7
FP	6 z 7	0 z 7
FN	3 zo 22	6 zo 22
Presnosť	0,76	1
Citlivosť	0,8636364	0,7272727
Úspešnosť	0,6896552	0,7931035

V tomto prípade môžeme vidieť, že lepšie výsledky vo všeobecnosti dosiahol algoritmus DT (CART), ale ak sa pozrieme na konkrétne triedy, zistíme, že Random

Forest je o niečo presnejší a stále platí, že je pre nás dôležitá hodnota TP, čo je počet korektne klasifikovaných podvrhnutých emailov. Nezáleží na tom, že algoritmus zle klasifikoval legitímne emaily, pretože administrátor stále môže zmeniť rozhodnutie systému. Horšie by bolo, ak by systém nekorektne klasifikoval viac podvrhnutých emailov. V tab. 9 a tab. 10 môžeme vidieť ako tieto algoritmy klasifikovali emaily:

Tab. 9 Vyhodnotenie algoritmu Random Forest pri testovaní systému

RF		**Predikované triedy**			
		legitímny	**phishing**	**scam**	**spam**
Očakávané triedy	**legitímny**	1	1	0	5
	phishing	0	3	0	2
	scam	2	2	4	3
	spam	2	0	0	4

Tab. 10 Vyhodnotenie algoritmu Decision Tree (CART) pri testovaní systému

DT (CART)		**Predikované triedy**			
		legitímny	**phishing**	**scam**	**spam**
Očakávané triedy	**legitímny**	7	0	0	0
	phishing	0	5	0	0
	scam	5	1	5	0
	spam	2	4	0	0

Výsledky týchto algoritmov sú veľmi podobné. Algoritmus RF zle klasifikoval len 4 podvrhnuté emaily a DT (CART) 7 emailov. A teda sa nám potvrdilo to, že sledovať prvotne výsledok algoritmu Random Forest je správny prístup. Vytvorili sme systém, ktorý dokáže pomôcť mnoho ľuďom na našej univerzite. Vďaka tomuto systému budú správcovia odľahčení od nadbytočnej práce a teda nebudú musieť riešiť problémy s podozrivými emailami. To za nich vyrieši náš systém, ktorý v prípade, že podobný email už registruje v databáze, hneď ho zaklasifikuje do konkrétnej triedy a automaticky odošle odpoveď nahlasovateľovi. Pomôže nielen správcom, ale tiež používateľom, ktorí si kedykoľvek budú môcť overiť každý email, ktorý vyzerá podozrivo.

Záver

Nepísaným cieľom tejto práce bolo odľahčiť správcov od nadbytočnej práce a tiež umožniť užívateľom overiť si, či emaily, ktoré dostávajú, sú bezpečné.

V prvej kapitole sme sa venovali sociálnemu inžinierstvu, jeho typom a tiež formám. Sociálne inžinierstvo môžeme rozdeliť na 2 typy a to na založené na ľuďoch a založené na technológiách. V našej práci sme sa zamerali na technológie. Foriem sme našli 5 – spam, scam, phishing, spearphishing a hoax, no obmedzili sme sa len na spam, scam a phishing, keďže zvyšné kategórie sa nám nepodarilo jednoznačne špecifikovať.

V druhej kapitole sme popísali niekoľko podobných prác a ich pohľad na problém detekcie podvodných emailov. Zistili sme, že keď aj niekto vytváral systém na detekciu, tak klasifikoval len na dve, poprípade tri triedy. Vo všeobecnosti sa ale väčšina zamerala na phishing a nonphishing emaily.

Predprípravná fáza bola v celku zložitá, pretože vo všeobecnosti neexistuje dataset, ktorý by mal priamo rozdelené emaily do kategórií. Spočiatku sme pracovali s datasetom rozdeleným mnou, ale neskôr sme požiadali ľudí z akademického CSIRT tímu o vyplnenie formulára na klasifikáciu emailov. Teda po niekoľkých týždňoch sme získali rozdelenie datasetu, nad ktorým sme mohli otestovať efektívnosť niekoľkých algoritmov dolovania dát, medzi ktoré patrí napríklad Random Forest alebo k-NN. Najlepšiu úspešnosť sme dosiahli pomocou algoritmu Random Forest pričom percentuálne dosiahol takmer 90%.

V rámci systému sme vytvorili Add-in do Outlooku, pomocou ktorého môže používateľ preposlať podozrivý email na server, tiež monitorujeme, či sa na spomínanom serveri nevyskytol nový email a ak áno, skontrolujeme či existuje podobný email – v prípade, že existuje, vieme ho hneď kategorizovať a odoslať používateľovi spätnú väzbu, ak nie, pomocou algoritmu Random Forest nanovo klasifikujeme email, pričom si pamätáme aj výsledky ostatných algoritmov.

Tiež sme vytvorili administrátorské rozhranie, vďaka ktorému má administrátor dohľad nad celým systémom a teda sme nezanedbali ľudský faktor. Dôležitou časťou posledného cieľa je otestovanie systému, ktoré sme vykonali so 44 emailami. Sledovali sme, či algoritmy korektne klasifikovali email ako legitímny alebo podvrhnutý a následne aj to, či email zaradil do správnej triedy.

Zoznam použitej literatúry

1. Projekt Statista [online]. [cit. 2019-04-19]. Dostupné z: https://www.statista.com/
2. PELTIER, Thomas R. Social engineering: concepts and solutions. Information Security Journal, 2006, 15.5: 13.
3. REYNOLDS, Vince. 2016. Social Engineering - The Art of Psychological Warfare, Human Hacking, Persuasion & Deception. Createspace, 2016. 106 s. ISBN 9781523850938
4. CHAUDHRY, Junaid Ahsenali; RITTENHOUSE, Robert G. Phishing: Classification and CounterMeasures. In: Multimedia, Computer Graphics and Broadcasting (MulGraB), 2015 7th International Conference on. IEEE, 2015. p. 28-31.
5. CAPUTO, Deanna D., et al. Going spear phishing: Exploring embedded training and awareness. IEEE Security & Privacy, 2014, 12.1: 28-38.
6. Method for recognizing spam email Blanzieri, E. and A. Bryl, A survey of learning-based techniques of email spam filtering. Artif. Intell. Rev., 2008. 29(1): p. 63-92.
7. ISACENKOVA, Jelena, et al. Inside the scam jungle: A closer look at 419 scam email operations. EURASIP Journal on Information Security, 2014, 2014.1: 4.
8. ISHAK, Adzlan; CHEN, Y. Y.; YONG, Suet-Peng. Distance-based hoax detection system. In: Computer & Information Science (ICCIS), 2012 International Conference on. IEEE, 2012. p. 215-220.
9. LIDDY, Elizabeth D. Natural language processing. 2001.
10. PENG, Tianrui; HARRIS, Ian; SAWA, Yuki. Detecting phishing attacks using natural language processing and machine learning. In: 2018 IEEE 12th International Conference on Semantic Computing (ICSC). IEEE, 2018. p. 300-301.
11. AGGARWAL, Shivam; KUMAR, Vishal; SUDARSAN, S. D. Identification and detection of phishing emails using natural language processing techniques. In: Proceedings of the 7th International Conference on Security of Information and Networks. ACM, 2014. p. 217.
12. Projekt WordNet [online]. [cit. 2019-04-19]. Dostupné z: https://wordnet.princeton.edu/

13. VERMA, Rakesh; SHASHIDHAR, Narasimha; HOSSAIN, Nabil. Detecting phishing emails the natural language way. In: European Symposium on Research in Computer Security. Springer, Berlin, Heidelberg, 2012. p. 824-841.
14. GASCON, Hugo, et al. Reading between the lines: content-agnostic detection of spear-phishing emails. In: International Symposium on Research in Attacks, Intrusions, and Defenses. Springer, Cham, 2018. p. 69-91.
15. SABERI, Alireza; VAHIDI, Mojtaba; BIDGOLI, Behrouz Minaei. Learn to detect phishing scams using learning and ensemble? methods. In: Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Workshops. IEEE Computer Society, 2007. p. 311-314.
16. MOKHTARI, Maryam, et al. A novel method in scam detection and prevention using data mining approaches. Proceedings of IDMC2008, 2008, 1-11.
17. WESTON, Jason; WATKINS, Chris. Multi-class support vector machines. Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May, 1998.
18. FORM, Lew May, et al. Phishing email detection technique by using hybrid features. In: 2015 9th International Conference on IT in Asia (CITA). IEEE, 2015. p. 1-5.
19. Projekt SpamAssassin [online]. [cit. 2019-04-19]. Dostupné z: http://spamassassin.apache.org/old/publiccorpus/
20. Projekt Nazario [online]. [cit. 2019-04-19]. Dostupné z: http://monkey.org/%7Ejose/wiki/doku.php?id=PhishingCorpus
21. BERGHOLZ, Andre, et al. Improved Phishing Detection using Model-Based Features. In: CEAS. 2008.
22. NIU, Weina, et al. Phishing Emails Detection Using CS-SVM. In: 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC). IEEE, 2017. p. 1054-1059.
23. SHARMA, Amit Kumar; PRAJAPAT, Sudesh Kumar; ASLAM, Mohammed. A Comparative Study Between Naive Bayes and Neural Network (MLP) Classifier for Spam Email Detection. In: IJCA Proceedings on National Seminar on Recent

Advances in Wireless Networks and Communications. Foundation of Computer Science (FCS). 2014. p. 12-16.

24. HSSINA, Badr, et al. A comparative study of decision tree ID3 and C4. 5. International Journal of Advanced Computer Science and Applications, 2014, 4.2: 0-0.
25. TOOLAN, Fergus; CARTHY, Joe. Feature selection for spam and phishing detection. In: 2010 eCrime Researchers Summit. IEEE, 2010. p. 1-12.
26. GENUER, Robin, et al. Random forests for big data. Big Data Research, 2017, 9: 28-46.
27. BREIMAN, Leo. Bagging predictors. Machine learning, 1996, 24.2: 123-140.
28. YASIN, Adwan; ABUHASAN, Abdelmunem. An intelligent classification model for phishing email detection. arXiv preprint arXiv:1608.02196, 2016.
29. FETTE, Ian; SADEH, Norman; TOMASIC, Anthony. Learning to detect phishing emails. In: Proceedings of the 16th international conference on World Wide Web. ACM, 2007. p. 649-656.
30. ABU-NIMEH, Saeed, et al. A comparison of machine learning techniques for phishing detection. In: Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit. ACM, 2007. p. 60-69.
31. HAMID, Isredza Rahmi A.; ABAWAJY, Jemal; KIM, Tai-hoon. Using feature selection and classification scheme for automating phishing email detection. Studies in informatics and control, 2013, 22.1: 61-70.
32. Projekt Enron-Spam [online]. [cit. 2019-04-19]. Dostupné z: http://www2.aueb.gr/users/ion/data/enron-spam/
33. Projekt Scikit-learn [online]. [cit. 2019-04-19]. Dostupné z: https://scikit-learn.org/stable/
34. Projekt Python [online]. [cit. 2019-04-19]. Dostupné z: https://www.python.org/
35. Projekt Outlook [online]. [cit. 2019-04-19]. Dostupné z: https://outlook.live.com/
36. Projekt Office365 [online]. [cit. 2019-04-19]. Dostupné z: https://www.office.com/
37. Projekt VisualStudio [online]. [cit. 2019-04-19]. Dostupné z: https://visualstudio.microsoft.com/

38. Projekt Fuzzywuzzy [online]. [cit. 2019-04-19]. Dostupné z: https://github.com/seatgeek/fuzzywuzzy
39. Projekt Levensthein [online]. [cit. 2019-04-19]. Dostupné z: http://www.levenshtein.net/
40. Projekt SsDeep [online]. [cit. 2019-04-19]. Dostupné z: https://ssdeep-project.github.io/ssdeep/index.html

Projekt MySQL [online]. [cit. 2019-04-19]. Dostupné z: https://www.mysql.com/

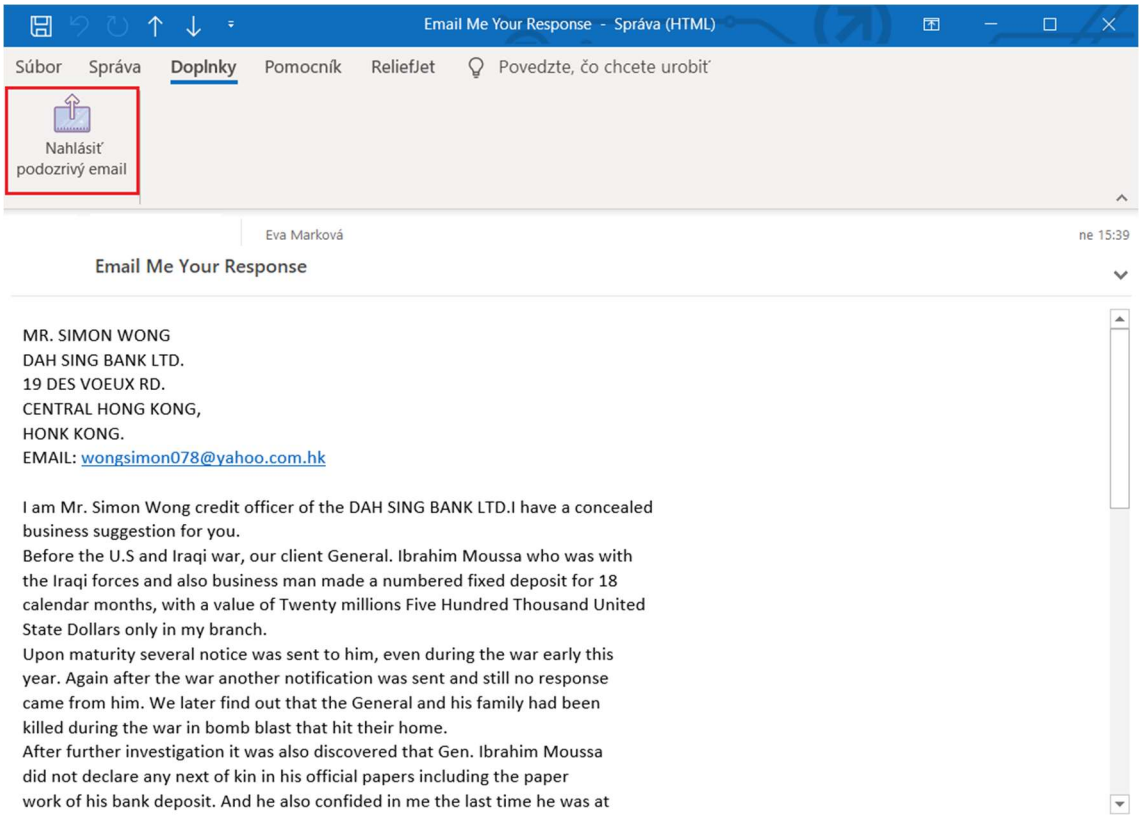
Prílohy

Príloha A: CD médium – bakalárska práca v elektronickej podobe, zdrojový kód systému na detekciu foriem sociálneho inžinierstva v emailovej komunikácii, manuál na použitie systému, dataset s 298 emailami

Príloha B: Manuál na použitie systému na detekciu sociálneho inžinierstva v emailovej komunikácii

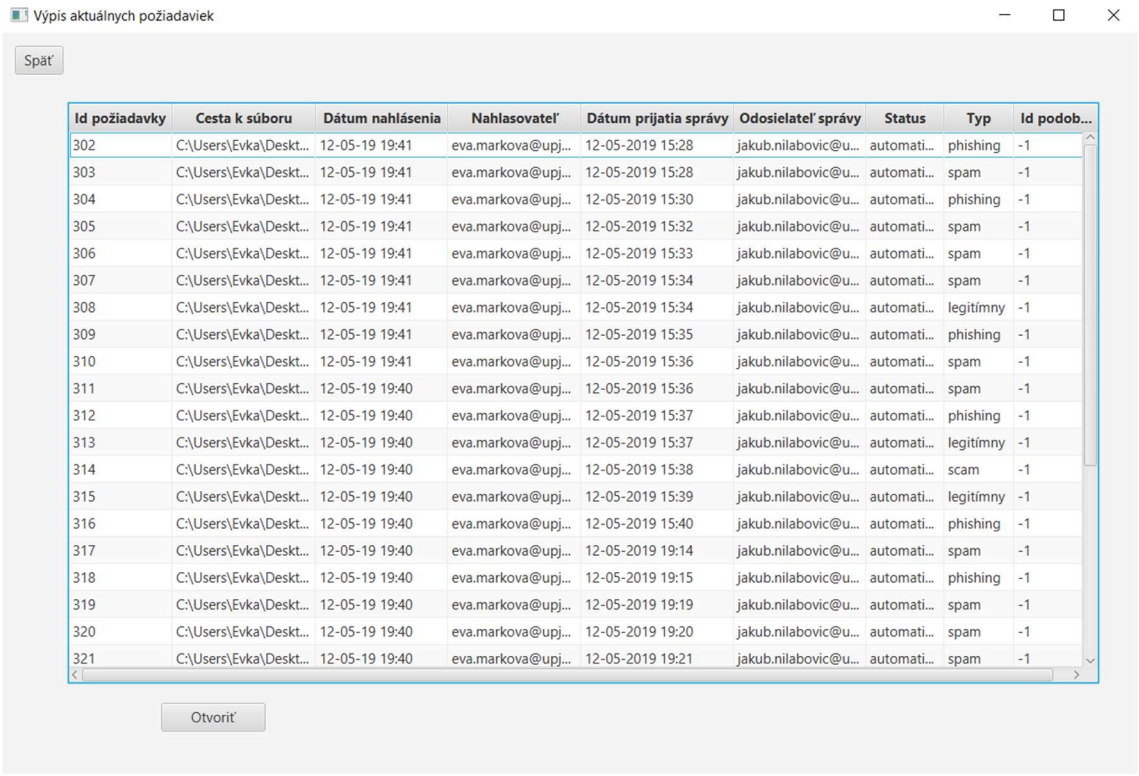
Príloha B: Manuál na použitie systému na detekciu sociálneho inžinierstva v emailovej komunikácii

V prípade, že vo svojej schránke natrafíte na email, ktorý vzbudzuje podozrenie, že sa jedná o podvrhnutý email, dvojklikom ho otvorte v novom okne. Po kliknutí na záložku „Doplnky“ uvidíte tlačidlo, ktoré je v červenom obdĺžniku na obr. 1. Po kliknutí naň sa email prepošle na server, kde sa skontroluje.



Obr. 1 Náhľad tlačidla na preposlanie podozrivého emailu

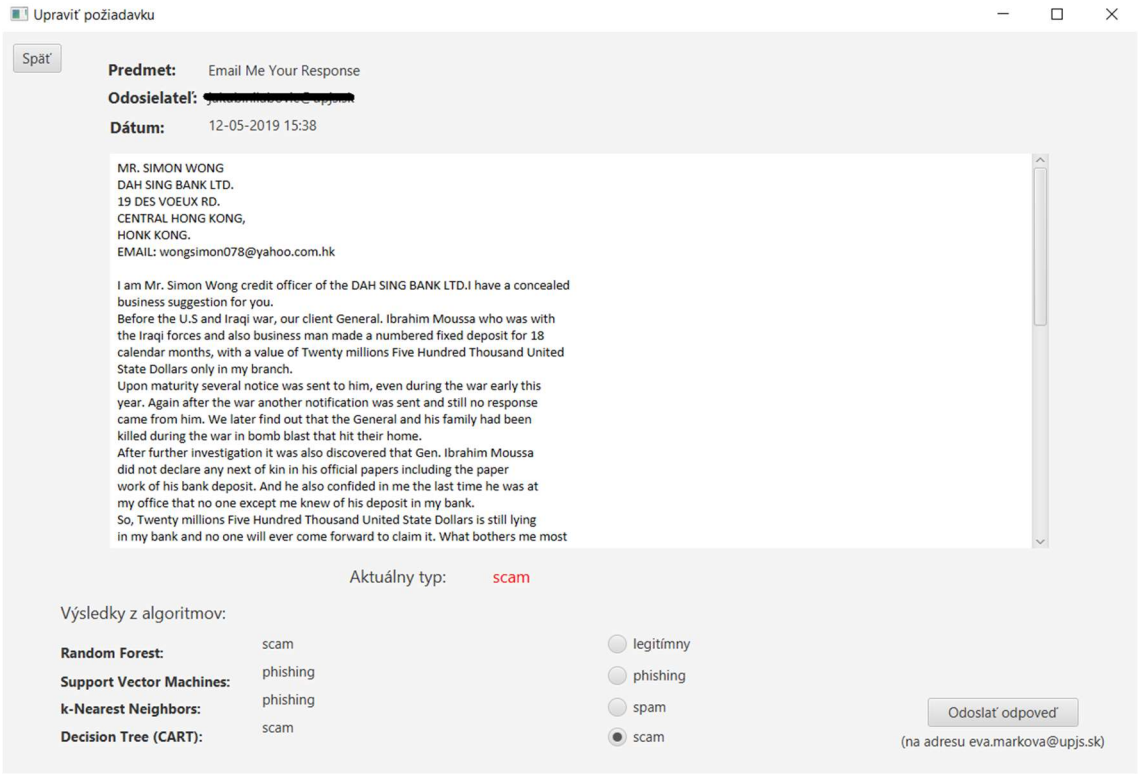
V administrátorskom rozhraní po kliknutí na tlačidlo „Požiadavky“ je vidieť rozhranie napríklad s nevyriešenými požiadavkami ako na obr. 2:



Id požiadavky	Cesta k súboru	Dátum nahlásenia	Nahlasovateľ	Dátum prijatia správy	Odosielateľ správy	Status	Typ	Id podob...
302	C:\Users\Evka\Deskt...	12-05-19 19:41	eva.markova@upj...	12-05-2019 15:28	jakub.nilabovic@u...	automati...	phishing	-1
303	C:\Users\Evka\Deskt...	12-05-19 19:41	eva.markova@upj...	12-05-2019 15:28	jakub.nilabovic@u...	automati...	spam	-1
304	C:\Users\Evka\Deskt...	12-05-19 19:41	eva.markova@upj...	12-05-2019 15:30	jakub.nilabovic@u...	automati...	phishing	-1
305	C:\Users\Evka\Deskt...	12-05-19 19:41	eva.markova@upj...	12-05-2019 15:32	jakub.nilabovic@u...	automati...	spam	-1
306	C:\Users\Evka\Deskt...	12-05-19 19:41	eva.markova@upj...	12-05-2019 15:33	jakub.nilabovic@u...	automati...	spam	-1
307	C:\Users\Evka\Deskt...	12-05-19 19:41	eva.markova@upj...	12-05-2019 15:34	jakub.nilabovic@u...	automati...	spam	-1
308	C:\Users\Evka\Deskt...	12-05-19 19:41	eva.markova@upj...	12-05-2019 15:34	jakub.nilabovic@u...	automati...	legitímny	-1
309	C:\Users\Evka\Deskt...	12-05-19 19:41	eva.markova@upj...	12-05-2019 15:35	jakub.nilabovic@u...	automati...	phishing	-1
310	C:\Users\Evka\Deskt...	12-05-19 19:41	eva.markova@upj...	12-05-2019 15:36	jakub.nilabovic@u...	automati...	spam	-1
311	C:\Users\Evka\Deskt...	12-05-19 19:40	eva.markova@upj...	12-05-2019 15:36	jakub.nilabovic@u...	automati...	spam	-1
312	C:\Users\Evka\Deskt...	12-05-19 19:40	eva.markova@upj...	12-05-2019 15:37	jakub.nilabovic@u...	automati...	phishing	-1
313	C:\Users\Evka\Deskt...	12-05-19 19:40	eva.markova@upj...	12-05-2019 15:37	jakub.nilabovic@u...	automati...	legitímny	-1
314	C:\Users\Evka\Deskt...	12-05-19 19:40	eva.markova@upj...	12-05-2019 15:38	jakub.nilabovic@u...	automati...	scam	-1
315	C:\Users\Evka\Deskt...	12-05-19 19:40	eva.markova@upj...	12-05-2019 15:39	jakub.nilabovic@u...	automati...	legitímny	-1
316	C:\Users\Evka\Deskt...	12-05-19 19:40	eva.markova@upj...	12-05-2019 15:40	jakub.nilabovic@u...	automati...	phishing	-1
317	C:\Users\Evka\Deskt...	12-05-19 19:40	eva.markova@upj...	12-05-2019 19:14	jakub.nilabovic@u...	automati...	spam	-1
318	C:\Users\Evka\Deskt...	12-05-19 19:40	eva.markova@upj...	12-05-2019 19:15	jakub.nilabovic@u...	automati...	phishing	-1
319	C:\Users\Evka\Deskt...	12-05-19 19:40	eva.markova@upj...	12-05-2019 19:19	jakub.nilabovic@u...	automati...	spam	-1
320	C:\Users\Evka\Deskt...	12-05-19 19:40	eva.markova@upj...	12-05-2019 19:20	jakub.nilabovic@u...	automati...	spam	-1
321	C:\Users\Evka\Deskt...	12-05-19 19:40	eva.markova@upj...	12-05-2019 19:21	jakub.nilabovic@u...	automati...	spam	-1

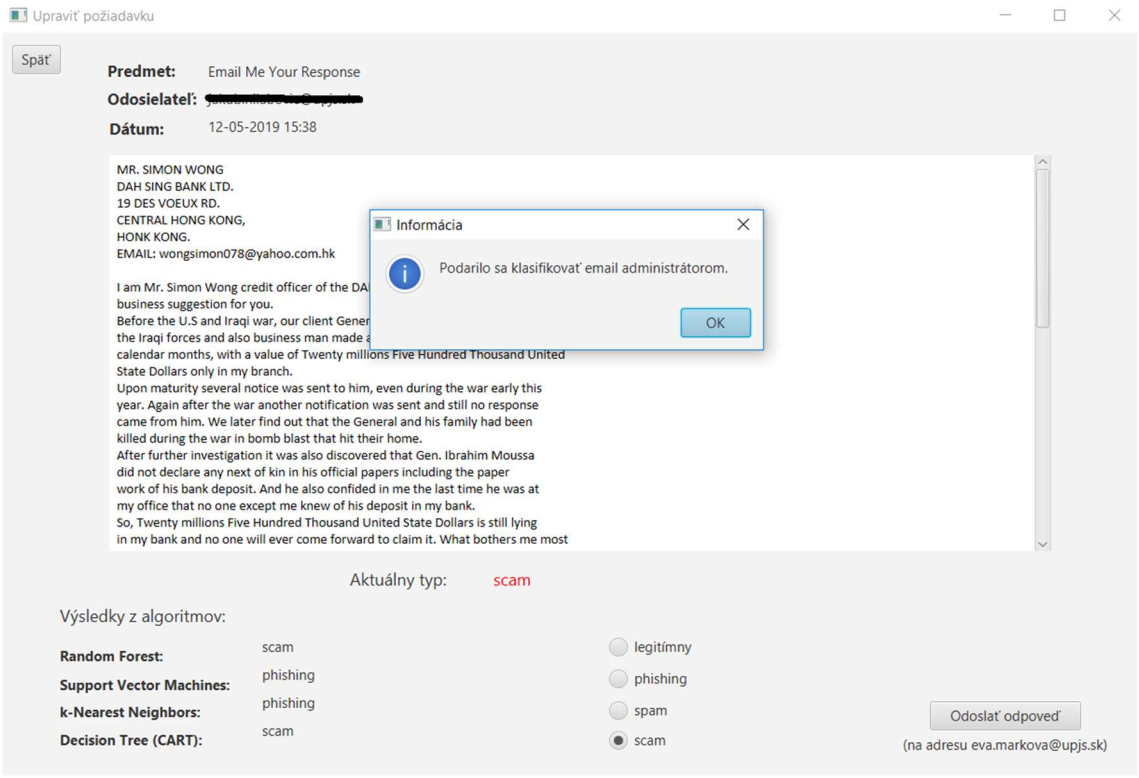
Obr. 2 Zoznam požiadaviek v administrátorskom rozhraní

V prípade, že prichádza mnoho požiadaviek s podobnými emailami, ktoré evidujeme v databáze, systém automaticky odpovedá používateľovi na základe spomínaného podobného emailu. Dôležité sú nevyriešené požiadavky – z nich si administrátor vyberie jednu. Na obr. 3 je email s id = 314.



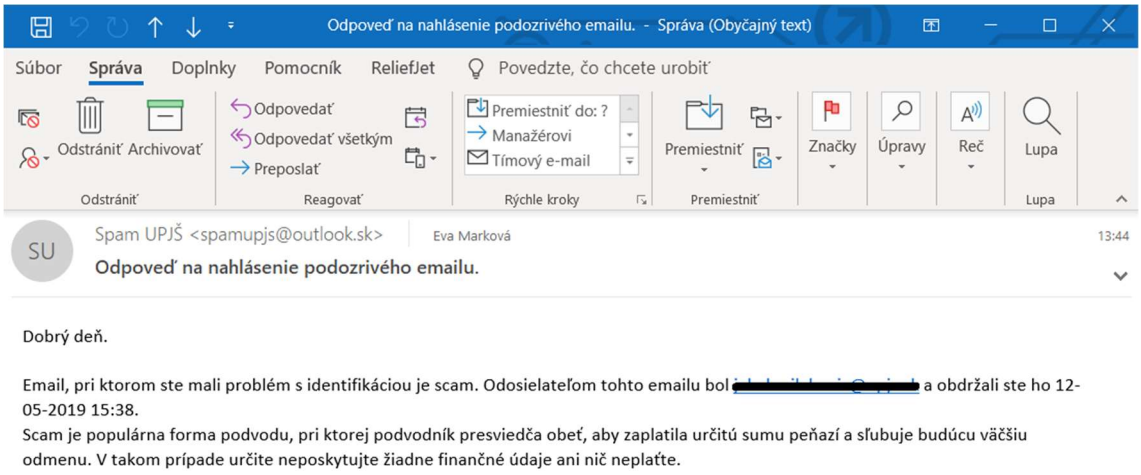
Obr. 3 Náhľad emailu s id = 314 v administrátorskom rozhraní

Na obr. 3 tiež môžeme vidieť ako v prípade tohto emailu rozhodli algoritmy, kto je odosielateľ, predmet emailu a podobne. Administrátor si vyberie jednu z tried, do ktorej zaradí email a klikne na tlačidlo „Odoslať odpoveď“.



Obr. 4 Náhľad po kliknutí na tlačidlo "Odoslať odpoveď"

Po kliknutí na tlačidlo požiadavku považujeme za vyriešenú. Na obr. 4 vidíme vyskakovacie okno, ktoré nás informuje o tom, že sa podarilo klasifikovať email. Nahlasovateľovi sa odošle odpoveď v závislosti od triedy, ktorú administrátor vybral. V tomto prípade sme vyhodnotili, že sa jedná o scam a teda používateľ dostal odpoveď, ktorú vidíme na obr. 5:



Obr. 5 Odpoveď užívateľovi v prípade, že sa jedná o scam