Univerzita P. J. Šafárika v Košiciach
Prírodovedecká fakulta

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Diana Fortunová
Študijný program: informatika (jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: Informatika
Typ záverečnej práce: Diplomová práca
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Návrh konverzačného robota pre jeho použitie proti podvodnej emailovej komunikácii

Názov EN: Design of a conversational robot for its use against fraudulent e-mail communication

Cieľ:
1. Spracovať prehľad vyžitia konverzačných robotov v oblasti podvodnej emailovej komunikácie.
2. Porovnať existujúce prístupy a nástroje k automatizovanej komunikácii s podvodníkmi.
3. Navrhnúť a implementovať model pre emailovú komunikáciu s podvodníkmi.
4. Navrhnúť, implementovať časť systému konverzačného robota pre podvodnú emailovú komunikáciu, vyhodnotiť efektívnosť tohto systému.

Literatúra:
1) Jiao, A. (2020). An intelligent chatbot system based on entity extraction using RASA NLU and neural network. In Journal of Physics: Conference Series (Vol. 1487, No. 1, p. 012014). IOP Publishing.
2) Adamopoulou, E., & Moussiades, L. (2020, June). An overview of chatbot technology. In IFIP International Conference on Artificial Intelligence Applications and Innovations (pp. 373-383). Springer, Cham.
3) Abdul-Kader, S. A., & Woods, J. C. (2015). Survey on chatbot design techniques in speech conversation systems. International Journal of Advanced Computer Science and Applications, 6(7).
4) Sahin, M., Relieu, M., & Francillon, A. (2017). Using chatbots against voice spam: Analyzing {Lenny's} effectiveness. In Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017) (pp. 319-337).

Vedúci: doc. RNDr. Ľubomír Antoni, PhD.
Ústav : ÚINF - Ústav informatiky
Riaditeľ ústavu: doc. RNDr. Ondrej Krídlo, PhD.

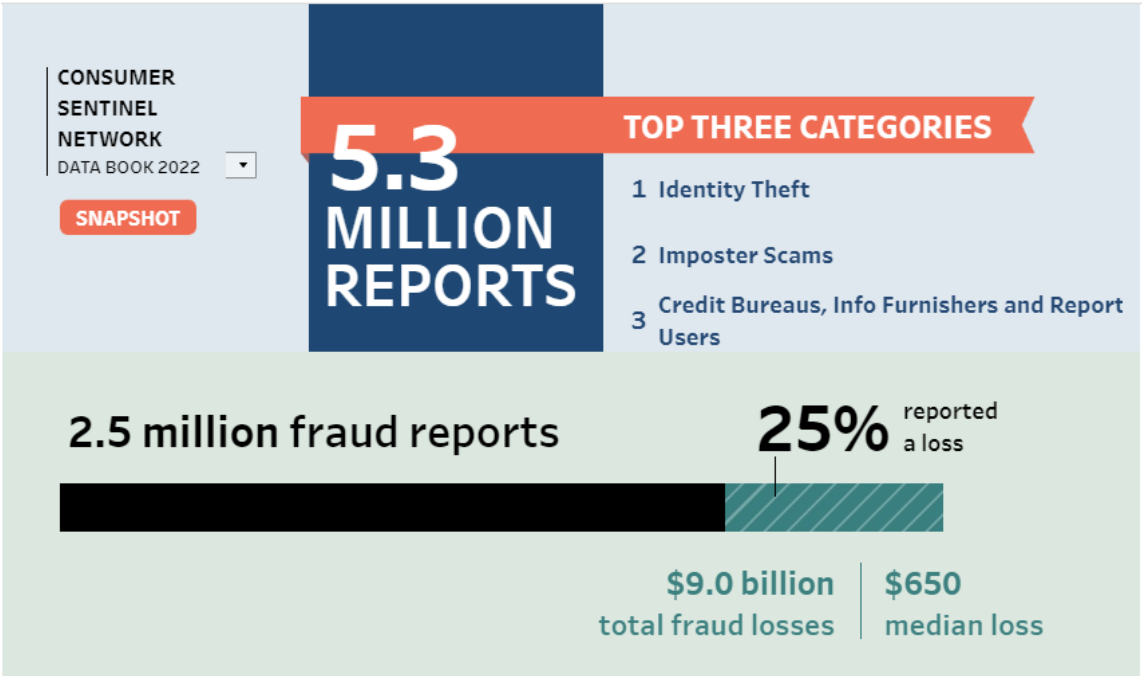
Dátum schválenia:

1. Úvod

S podvodnými typmi emailu sa stretol pravdepodobne každý, aspoň raz v živote. Chorá dcéra, ktorá potrebuje operáciu, zosnulý klient z Toga s rovnakým priezviskom ako obeť, ktorý zhodou okolností nemá žiadnu rodinu, ktorá by po ňom dedila. Vdova, ktorá túži po láske alebo podnikateľka, ktorá má ponuku, ktorá sa len tak neodmieta. A mnoho ďalších.

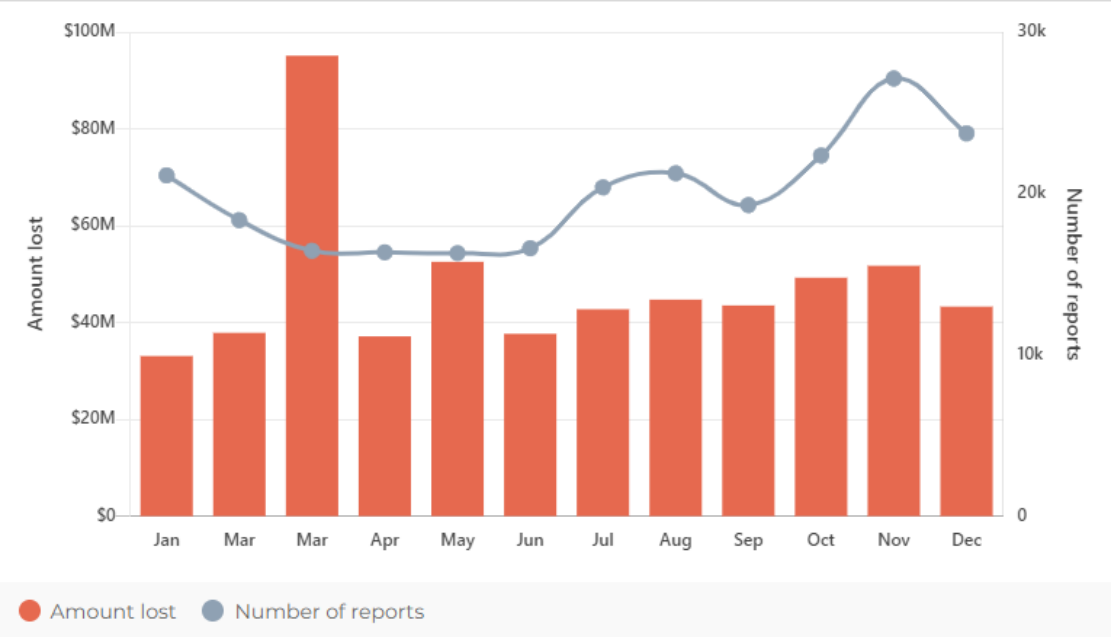
Domnienka, že táto forma podvodu je tu už dosť dlho na to, aby sa na nej niekto nachytal je však mylná. Podľa štatistík (Obr. 1) uvedených v knihe CONSUMER SENTINEL NETWORK spracovanej americkým FTC (Federal Trace Commission) z roku 2022 [1] (nové vydanie vychádza každý február) sa v rámci Spojených štátov počet nahlásení týchto podvodov ročne zvyšuje a zatiaľ nevyzerá, že by tento trend naberal inú trajektóriu.

Hlavným faktorom tohto nárastu je fakt, že podvodníci sa každým rokom zlepšujú a sú presvedčivejší, sofistikovanejší a aj jazyková bariéra sa pomaly stráca a chýb v daných správach je čoraz menej.



Obr. 1 FTC Consumer Sentinel Network DataBook 2022 [1]

Pre porovnanie spomeniem aj štatistiky zo SCAMWATCH [2], ktoré sledujú počet nahlásení podvodov sprostredkované austrálskym ACCC (Australian Competition & Customer Commission). Aj tieto štatistiky poukazujú na stúpajúcu tendenciu počtu nahlásení takéhoto typu podvodu:



Obr. 2 ACCC SCAMWATCH Statistics 2022 [2]

Čo sa týka rozdielov v rámci vekových kategórií, vieme s určitosťou povedať, že čo sa týka peňažných strát, ľudia v skupine nad 65 rokov majú o dosť väčší medián, ako mladšie ročníky. Taktiež je to skupina ľudí, ktorá tieto podvody len málokedy nahlási príslušným orgánom.

A keďže si tieto podvody ročne vyžiadajú niekoľko miliónov obetí, bolo by vhodné mať nejaký nástroj, ktorý by sa s týmto vedel vysporiadať. Riešením by mohol byť chatbot, inteligentný konverzačný robot, ktorý by s podvodníkom komunikoval namiesto obetí a zamestnal ho natoľko, aby sa nedostal k ďalšej osobe.

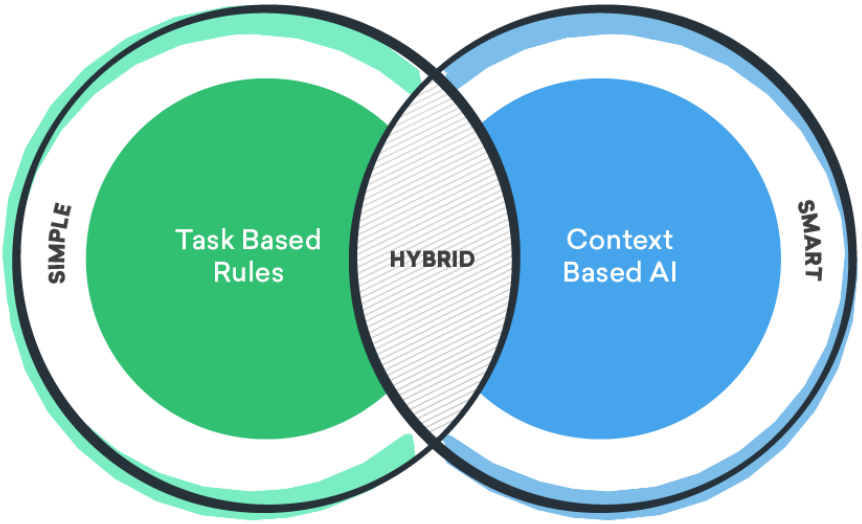
2. Chatbot

Chatbot, skratka pre chat robot alebo bot, je aplikácia, ktorá komunikuje s ľuďmi prostredníctvom príkazov používateľa. Tieto interakcie sa väčšinou uskutočňujú prostredníctvom hlasových a textových konverzácií. Je navrhnutý tak, aby replikoval vzor ľudskej interakcie, a tak umožnil ľudskú konverzáciu so strojmi.

2.1. Delenie chatbotov

Chatboty, poháňané AI, automatizovanými pravidlami, nástrojom na spracovanie prirodzeného jazyka a strojovým učením, spracúvajú údaje, aby doručili odpovede na požiadavky každého druhu.

Poznáme dva druhy chatbotov:



Obr. 3 Typy chatbotov [12]

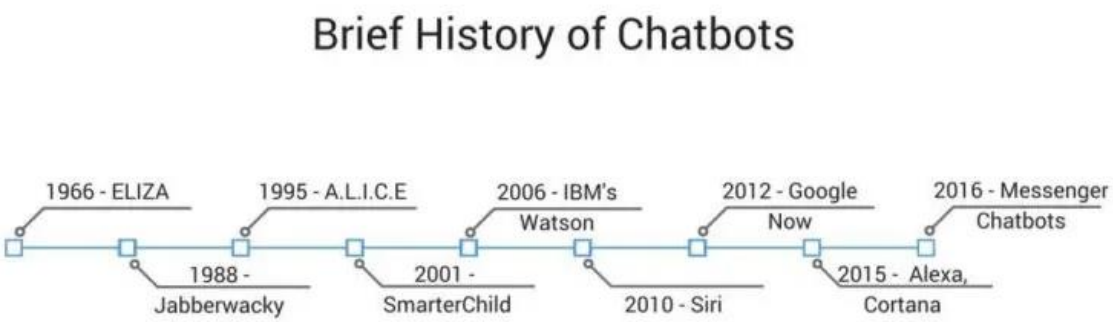
2.1.1. Úlohovo orientované (deklaratívne) chatboty

Sú to jednoúčelové programy, ktoré sa zameriavajú na vykonávanie jednej funkcie. Pomocou pravidiel, NLP a veľmi malého množstva ML generujú automatické, ale konverzačné odpovede na otázky používateľov. Interakcie s týmito chatbotmi sú vysoko špecifické a štruktúrované a sú najvhodnejšie pre podporné a servisné funkcie. Úlohovo orientované chatboty dokážu spracovať bežné otázky, ako sú otázky týkajúce sa pracovného času alebo jednoduchých transakcií, ktoré nezahŕňajú rôzne premenné. Aj keď používajú NLP, aby ich koncoví používatelia používali konverzačným spôsobom, ich schopnosti sú pomerne základné. Toto sú v súčasnosti najčastejšie používané chatboty.

2.1.2. Dátovo orientované a prediktívne (konverzačné) chatboty

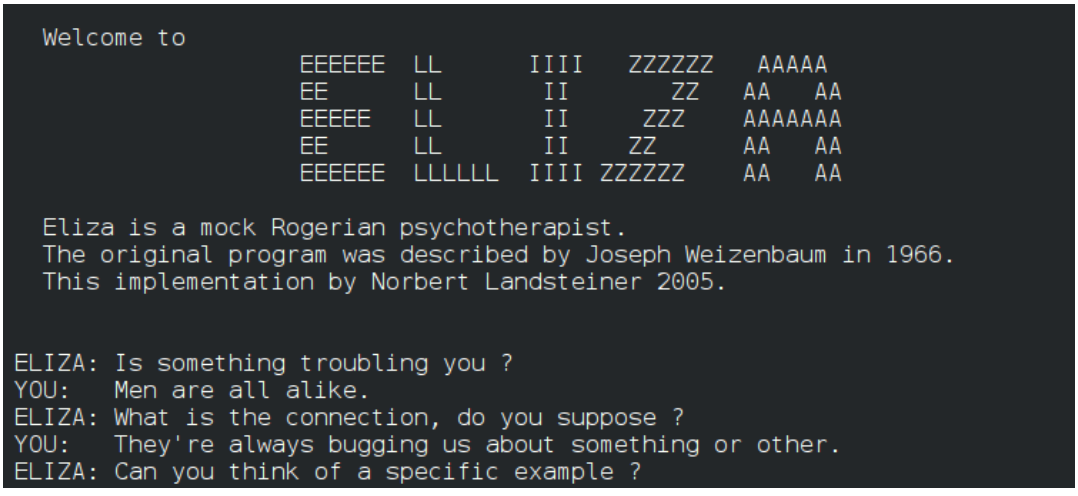
Sa často označujú ako virtuálni asistenti alebo digitálni asistenti a sú oveľa sofistikovanejšie, interaktívnejšie a personalizovanejšie ako chatboty orientované na úlohy. Tieto chatboty si uvedomujú kontext a využívajú porozumenie prirodzeného jazyka (NLU), NLP a ML, aby sa učili za chodu. Používajú predikčnú inteligenciu a analytiku, aby umožnili personalizáciu založenú na používateľských profiloch a správaní používateľov v minulosti. Digitálni asistenti sa môžu časom naučiť preferencie používateľa, poskytovať odporúčania a dokonca predvídať potreby. Okrem sledovania údajov a zámerov môžu iniciovať konverzácie. Apple Siri a Amazon Alexa sú príkladmi spotrebiteľsky orientovaných, dátovo orientovaných, prediktívnych chatbotov.

2.2. História chatbotov



Obr. 4 História chatbotov [13]

- ELIZA: Považuje sa za prvého chatbota v histórii počítačovej vedy, ktorého vyvinul Joseph Weizenbaum na Massachusettskom Inštitúte Techniky (MIT). Funguje tak, že rozpoznáva kľúčové slová alebo frázy zo vstupu a reprodukuje odpoveď pomocou týchto slov z vopred naprogramovaných odpovedí. Napríklad, ak človek povedal, že „Moja mama varí dobré jedlo“. ELIZA by zachytila slovo „matka“ a odpovedala by položením otvorenej otázky „Povedz mi viac o svojej rodine“. To vytvorilo ilúziu porozumenia a interakcie so skutočnou ľudskou bytosťou, hoci tento proces bol mechanizovaný.



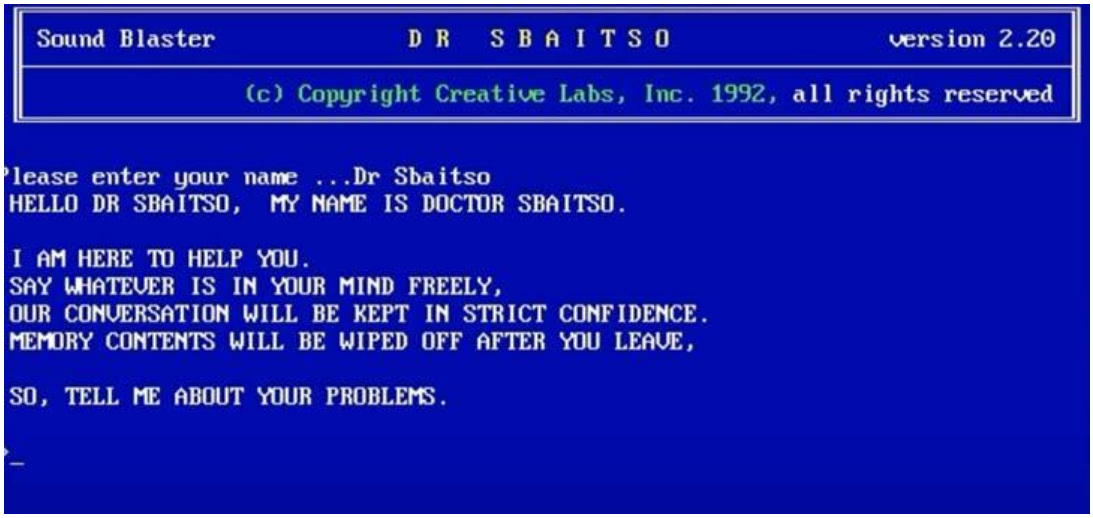
Obr. 5 Eliza [14]

- ALICE: Vyvinul ho v roku 1995 Richard Wallace. Na rozdiel od Elizy, chatbot ALICE dokázal využiť spracovanie prirodzeného jazyka, čo umožnilo sofistikovanejšiu konverzáciu. Revolučnosť spočívala v tom, že išlo o open-source.

- PARRY: Skonštruoval americký psychiater Kenneth Colby v roku 1972. Program napodobňoval pacienta so schizofréniou. Pokúša sa simulovať chorobu. Ide o program prirodzeného jazyka, ktorý pripomína myslenie jednotlivca. PARRY funguje prostredníctvom komplikovaného systému predpokladov, atribútov a „emocionálnych reakcií" spúšťaných zmenou váh priradených verbálnym vstupom. Na overenie práce bol PARRY testovaný pomocou variácie Turingovho testu. Bolo to na začiatku sedemdesiatych rokov, keď ľudskí vyšetrovatelia, ktorí interagovali s programom prostredníctvom vzdialenej klávesnice, boli slabí s viac než náhodnou presnosťou, aby odlíšili PARRYho od pôvodného nerozumného jednotlivca.

- JABBERWACKY: Jabberwacky je chatterbot vytvorený britským programátorom Rollom Carpenterom. Jeho cieľom je „simulovať prirodzený ľudský rozhovor zaujímavým, zábavným a vtipným spôsobom". Ide o skorý pokus o vytvorenie umelej inteligencie prostredníctvom ľudskej interakcie. Stanoveným účelom projektu bolo vytvoriť umelú inteligenciu, ktorá je schopná prejsť Turingovým testom. Je navrhnutý tak, aby napodobňoval ľudskú interakciu a viedol rozhovory s používateľmi. Nie je určený na vykonávanie žiadnych iných funkcií.

- DR. SBAITSO: Dr. Sbaitso je chatbot vytvorený spoločnosťou Creative Labs pre MS-Dos v roku 1992. Je to jeden z prvých pokusov o začlenenie A.I. do chatbota a je uznávaný pre svoj plne hlasový chatovací program. Program by s používateľom konverzoval, ako keby to bol psychológ. Väčšina jeho odpovedí bola v zmysle „Prečo sa tak cítite?" namiesto akejkoľvek komplikovanej interakcie.



Obr. 6 Dr. Sbaitso [15]

- MITSUKU: Kuki je chatbot vytvorený pomocou technológie AIML (predstavenej u ALICE) od Steva Worswicka. Tvrdí, že ide o 18-ročnú chatbotku z anglického Leedsu. Obsahuje všetky súbory Alice AIML s mnohými doplnkami z konverzácií generovaných používateľmi a vždy sa na nej pracuje. Jej inteligencia zahŕňa schopnosť uvažovať o daných objektoch. Napríklad, ak sa niekto spýta „Môžem zjesť dom?“, Kuki vyhľadá vlastnosti pre „dom“. Nájde hodnotu „made_from“ nastavenú na „brick“ a odpovie „nie“, pretože dom nie je jedlý. Môže hrať hry a robiť magické triky na žiadosť používateľa. V roku 2015 konverzovala v priemere viac ako štvrť milióna krát denne.

- SIRI: Siri bola vytvorená spoločnosťou Apple pre iOS v roku 2010; je to inteligentný osobný asistent a vzdelávací navigátor, ktorý používa používateľské rozhranie prirodzeného jazyka. Potom to pripravilo systém pre všetkých robotov AI a PA.

- GOOGLE: Asistent Google bol spustený na Google Inch v roku 2012. Odpovedá na otázky, vykonáva akcie prostredníctvom požiadaviek odoslaných súborom webových služieb a poskytuje odporúčania. Bol súčasťou balíka aktualizácií a úprav používateľského rozhrania pre mobilné vyhľadávanie, ktorý zahŕňal prenosnú asistentku so ženským hlasom, ktorá konkuruje Siri od Apple.

- CORTANA: Cortana bola prvýkrát predstavená na vývojárskej konferencii spoločnosti Microsoft Build 2014 a stala sa priamo integrovanou do zariadení Windows Phone a Windows 10 PC. Tento program využíva rozpoznávanie hlasu a príslušné algoritmy na získavanie a reagovanie na hlasové príkazy. Cortana môže vykonávať úlohy, ako sú pripomienky na základe času, miest alebo ľudí, odosielať e-maily a texty, vytvárať a spravovať zoznamy, chatovať a hrať hry, okrem iného nájsť fakty, súbory, miesta a informácie.

- ALEXA: Alexa je inteligentný osobný asistent vyvinutý spoločnosťou Amazon. Bola predstavená v roku 2014 a teraz je zabudovaná do zariadení, ako sú Amazon Echo, Echo Dot, Echo Show a ďalšie. K dispozícii je tiež aplikácia Alexa a ďalšie zariadenia od výrobcov tretích strán, ktoré majú v sebe zabudovanú Alexu.

- CHATGPT: ChatGPT je rozsiahly jazykový model vyškolený OpenAI. Bol založený tímom OpenAI v roku 2021. Je navrhnutý tak, aby pomáhal používateľom pri vytváraní ľudského textu na základe daného vstupu. ChatGPT možno použiť na rôzne úlohy vrátane generovania konverzácií a prekladu jazyka. Model je trénovaný na veľkom množstve údajov, čo mu umožňuje generovať text, ktorý je často ťažké rozlíšiť od textu napísaného človekom. ChatGPT bol chválený pre svoju schopnosť vytvárať prirodzene znejúci text a jeho potenciálne aplikácie v rôznych oblastiach.

3. Bezpečnostné riziká chatbotov

Existujú obavy o bezpečnosť systémov chatbotov. Chatboty sú často pripojené k internetu, čo znamená, že sú zraniteľné voči hackerom a kybernetickým útokom. Ak je chatbot systém napadnutý, môžu byť odcudzené osobné informácie.

Ďalším problémom je nedostatočná transparentnosť v tom, ako sa údaje chatbotov zhromažďujú, uchovávajú a ako sa k nim pristupuje. Mnoho vývojárov chatbotov jasne nevysvetľuje, ako zhromažďujú a používajú údaje o používateľoch, takže používatelia nevedia, ako sa s ich osobnými údajmi nakladá. Tento nedostatok transparentnosti môže viesť k nedôvere voči chatbotom a neochote ich používať.

Okrem toho existuje obava, že údaje chatbota môžu byť zdieľané s tretími stranami bez vedomia alebo súhlasu používateľa. To by mohlo potenciálne viesť k tomu, že údaje sa použijú na cielenú reklamu alebo iné účely, s ktorými používateľ nemusí súhlasiť.

Používatelia tiež zohrávajú úlohu pri ochrane svojho súkromia pri používaní chatbotov. Vždy sa odporúča poskytnúť minimálne potrebné informácie, ktoré neobsahujú osobné údaje.

Dôkazom o obavách o je aj fakt, že taliansky orgán pre ochranu súkromia pred niekoľkými mesiacmi zakázal používanie ChatGPT z dôvodu narušenia bezpečnosti a domnienke, že populárny chatbot používa osobné údaje používateľov na svoje trénovanie. (ChatGPT v súčasnosti opäť beží na talianskom území)

3.1. Data poisoning

Je novo koncipovaný kybernetický útok, ktorý sa priamo zameriava na umelú inteligenciu. Technológia AI sa učí z datasetov a používa tieto informácie na vykonávanie úloh. To platí pre všetky programy AI bez ohľadu na ich účel alebo funkciu.

Kybernetickí útočníci môžu nájsť spôsoby, ako zasahovať do daných datasetov používaných na trénovanie AI, čo im umožňuje manipulovať s ich rozhodnutiami a reakciami. Umelá inteligencia použije informácie zo zmenených údajov a vykoná akcie, ktoré útočníci chcú. Keď AI začne získavať otrávené údaje, je ťažké ich odhaliť a môže to viesť k významnému narušeniu kybernetickej bezpečnosti, ktoré zostane dlho nepovšimnuté.

3.2. Škodlivé predsudky

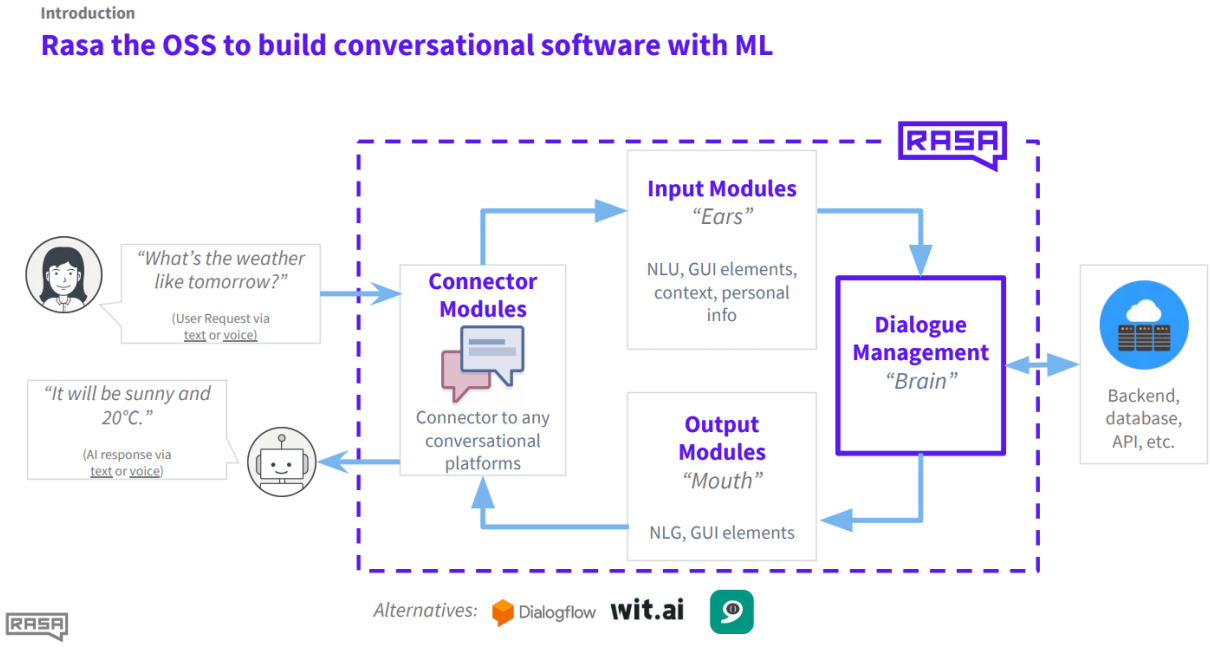
Jedným z najväčších nebezpečenstiev chatbotov AI je ich tendencia k škodlivým predsudkom. Keďže umelá inteligencia vytvára spojenia medzi dátovými bodmi, ktoré ľuďom často unikajú, dokáže vo svojich tréningových dátach zachytiť jemné, implicitné predsudky, aby sa naučila byť diskriminačná. Výsledkom je, že chatboti sa môžu rýchlo naučiť chrliť rasistický, sexistický alebo inak diskriminačný obsah, aj keď v údajoch o tréningu nebolo nič také extrémne.

Najlepším príkladom je zrušený náborový robot Amazonu. V roku 2018 sa ukázalo, že Amazon opustil projekt AI, ktorého cieľom bolo predbežne posúdiť životopisy žiadateľov, pretože penalizoval žiadosti od žien. Keďže väčšina životopisov, na ktorých bot trénoval, bola mužská, naučil sa, že preferovaní sú mužskí žiadatelia, aj keď to v tréningových údajoch nie je výslovne uvedené.

Chatboty, ktoré používajú internetový obsah, aby sa naučili prirodzene komunikovať, majú tendenciu predviesť ešte extrémnejšie predsudky. V roku 2016 Microsoft predstavil chatbota s názvom Tay, ktorý sa naučil napodobňovať príspevky na sociálnych sieťach. V priebehu niekoľkých hodín začala tweetovať vysoko urážlivý obsah, čo viedlo Microsoft k pozastaveniu účtu čoskoro.

4. Náznak postupu riešenia a čiastkové ciele

Medzi základné vlastnosti konverzačných robotov patrí prijatie správy od ľudského používateľa, nájdenie kontextu tejto správy, jeho porozumenie, spracovanie a následné vygenerovanie odpovede v súlade s požiadavkami používateľa. Postup daných činností, je znázornený na obrázku 3:



Obr. 7 Moduly systému konverzačného robota [3]

Tieto vlastnosti vieme rozdeliť do niekoľkých modulov a to:

- Integračný modul virtuálneho asistenta pre rôzne platformy ktorého cieľom je prijímať dopyt od používateľa a následne odosielať vygenerovanú odpoveď,

- Vstupný modul tzv. uši chatbota, ktorý má za úlohu porozumieť prirodzenému jazyku, vyčítať kontext dopytu, určiť zámer daného používateľa a rozpoznať entity danej správy,

- Modul riadenia dialógu s používateľom, slúžiaci na monitorovanie aktuálneho stavu konverzácie a komunikáciu s databázou alebo rozhraním API v prípade potreby získania ďalších informácií do odpovede,

- Výstupný modul – ústa, slúžiace na generovanie prirodzeného jazyka, pre vygenerovanie adekvátnej odpovede v ľudskej reči pre daného používateľa.

S postupnou implementáciou strojového učenia sa jadrom systému inteligentných konverzačných robotov stáva technika spracovania prirodzeného jazyka (NLP), ktorá pozostáva z modulu na porozumenie prirodzenému jazyku (NLU) a generovanie prirodzeného jazyka (NLG), ktoré dokážu simulovať zmysluplný dialóg s človekom [4].

4.1. Klasifikácia entít z oblasti kybernetickej bezpečnosti

Správy prijaté konverzačným robotom obsahujú zámer používateľa. Tento zámer nám mnohokrát vedia priblížiť kľúčové slová, ktoré v našom kontexte budeme nazývať entity. Entity predstavujú vopred určené kategórie skupín objektov, ktoré majú význam a pochádzajú z prirodzeného jazyka. Konverzačné roboty však musia zachytávať vlastnú sadu kategórií entít podľa ich domény - oblasti záujmu (napr. poštový asistent, asistent online bankingu, ...)

Named entity recognition (NER) je pravdepodobne prvým krokom k extrakcii informácií. Snaží sa nájsť a klasifikovať pomenované entity v texte do vopred definovaných kategórií, ako sú mená osôb, organizácií , miesta, vyjadrenia časov, množstiev, peňažných hodnôt, percent atď. NER sa používa v mnohých oblastiach spracovania prirodzeného jazyka (NLP) a môže pomôcť odpovedať na mnohé otázky, ako napríklad:

- Ktoré spoločnosti boli spomenuté v spravodajskom článku ?
- Na aký produkt sa vzťahuje reklamácia od daného kupujúceho ?
- Akú diagnózu má daný pacient ?
- Aká je adresa, na ktorú je potrebné doručiť danú objednávku ?

Jedným z dostupných nástrojov, ktorý veľmi šikovne rieši danú problematiku napríklad v medicínskej sfére je **PoolParty** [5]. Založený na znalostnom grafe a algoritmoch strojového učenia NER. Tento nástroj umožňuje rozpoznať:

- Pomenované entity extrahované prostredníctvom NER na základe algoritmov strojového učenia.
- Koncepty extrahované pomocou extrakcie konceptov na základe PoolParty Extractor-a.
- Dôležité pojmy identifikované prostredníctvom štatistickej textovej analýzy.
- Typ dokumentu identifikovaný klasifikačnými algoritmami založenými na Machine Learning.

4.1.1. Ako na implementáciu ?

Klasifikáciu entít vieme dosiahnuť pomocou nástroja NLTK (Natural Language Toolkit) čo je popredná platforma pre vytváranie Python-ovských programov pre prácu s ľudskou rečou. Poskytuje ľahko použiteľné rozhrania pre viac ako 50 korpusov a lexikálnych zdrojov, ako je WordNet, spolu so sadou knižníc na spracovanie textu pre klasifikáciu, tokenizáciu, odvodzovanie, označovanie, analýzu a sémantické uvažovanie [6].

Tokenizácia je proces rozdelenia daného textu na jednotky nazývané tokeny. Tokeny

Identify named entities:

```
>>> entities = nltk.chunk.ne_chunk(tagged)
>>> entities
Tree('S', [('At', 'IN'), ('eight', 'CD'), ("o'clock", 'JJ'),
          ('on', 'IN'), ('Thursday', 'NNP'), ('morning', 'NN'),
      Tree('PERSON', [('Arthur', 'NNP')]),
          ('did', 'VBD'), ("n't", 'RB'), ('feel', 'VB'),
          ('very', 'RB'), ('good', 'JJ'), ('.', '.')])
```

Obr. 8 Jednoduchá identifikácia entít

môžu byť jednotlivé slová, frázy alebo dokonca celé vety. V procese tokenizácie môžu byť niektoré znaky, ako sú interpunkčné znamienka, vyradené. Tokeny sa zvyčajne stávajú vstupom pre procesy, ako je analýza a dolovanie textu [7]:

Tokenize and tag some text:

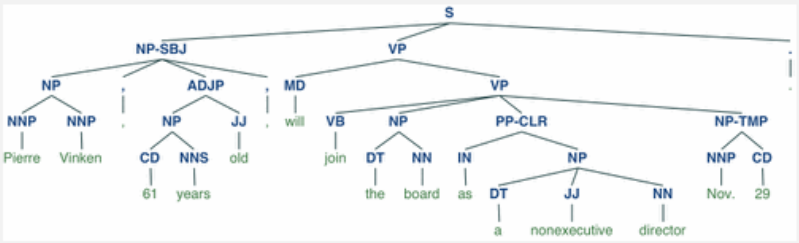
```
>>> import nltk
>>> sentence = """At eight o'clock on Thursday morning
... Arthur didn't feel very good."""
>>> tokens = nltk.word_tokenize(sentence)
>>> tokens
['At', 'eight', "o'clock", 'on', 'Thursday', 'morning',
'Arthur', 'did', "n't", 'feel', 'very', 'good', '.']
>>> tagged = nltk.pos_tag(tokens)
>>> tagged[0:6]
[('At', 'IN'), ('eight', 'CD'), ("o'clock", 'JJ'), ('on', 'IN'),
('Thursday', 'NNP'), ('morning', 'NN')]
```

Obr. 9 Tokenizácia textu pomocou NLTK

NLTK taktiež umožňuje tzv. Chunking, proces spracovania prirodzeného jazyka používaný na identifikáciu slovných druhov ako napríklad podstatných mien alebo krátkych fráz prítomných v danej vete. Výstup tejto metódy je taktiež možné zobraziť aj vo forme stromu:

Display a parse tree:

```
>>> from nltk.corpus import treebank
>>> t = treebank.parsed_sents('wsj_0001.mrg')[0]
>>> t.draw()
```



Obr. 10 Chunking tree

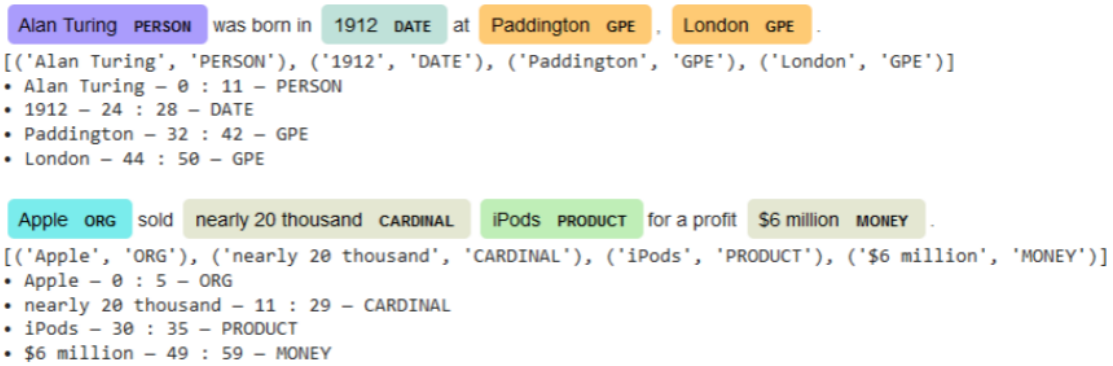
Ďalším nástrojom, umožňujúcim klasifikáciu entít v jazyku Python je bezplatná knižnica spaCy [8]. Hoci nepodporuje slovenský jazyk, v súčasnosti poskytuje jazykové modely pre viac ako 20 jazykov a umožňuje vytvorenie viacjazyčného modelu. Podobne ako NLTK taktiež disponuje možnosťou tokenizácie textu a hľadá medzi nimi súvis. Dokáže vykresliť vizualizáciu vzťahov medzi jednotlivými tokenmi, ale aj určiť ich slovné druhy alebo vypočítať podobnosť slov na základe vektorovej reprezentácie.

SpaCy bola trénovaná na korpuse OntoNotes 5 (rozsiahly korpus obsahujúci rôzne žánre textu (správy, konverzácie, weblogy, diskusné fóra, relácie,...) v troch jazykoch (angličtina, čínština a arabčina) a podporuje nasledujúce entity:

PERSON - People, including fictional.
NORP - Nationalities or religious or political groups.
FAC - Buildings, airports, highways, bridges, etc.
ORG - Companies, agencies, institutions, etc.
GPE - Countries, cities, states.
LOC - Non-GPE locations, mountain ranges, bodies of water.
PRODUCT - Objects, vehicles, foods, etc. (Not services.)
EVENT - Named hurricanes, battles, wars, sports events, etc.
WORK_OF_ART - Titles of books, songs, etc.
LAW - Named documents made into laws.
LANGUAGE - Any named language.
DATE - Absolute or relative dates or periods.
TIME - Times smaller than a day.
PERCENT - Percentage, including "%".
MONEY - Monetary values, including unit.
QUANTITY - Measurements, as of weight or distance.
ORDINAL - "first", "second", etc.
CARDINAL - Numerals that do not fall under another type.

Tieto entity by v komunikácii chatbota s ľuďmi s vysokou pravdepodobnosťou predstavovali kľúčové informácie. Množstvo z týchto entít je taktiež použiteľných pri analýze podvodných správ na zistenie napríklad požadovanej sumy, ktorú má obeť zaplatiť alebo deadline platby.

Okrem tejto funkcie spaCy poskytuje aj iné možnosti analýzy a spracovania textu v prirodzenom jazyku. Na úvod je potrebné zvoliť jazykový model, s ktorým chceme pracovať. Následne načítame textový reťazec, ktorý sa použitím funkcie na tokenizáciu konvertuje na objekt typu Doc, s ktorým knižnica spaCy ďalej pracuje. V ďalšom kroku je možné pre každú entitu z tohto objektu vypísať rôzne atribúty, v našom prípade máme na výstupe text nájdenej entity – text, začiatočnú a koncovú pozíciu jej znakov v texte – start_char i end_char a kategóriu, ktorá jej bola priradená – label_:



Obr. 11 Rozpoznávanie entít pomocou spaCy [4]

4.2. Textová analýza pomocou metód strojového učenia

Číselná reprezentácia textových dokumentov je náročná úloha v strojovom učení. Takáto reprezentácia môže byť použitá na mnohé účely, napríklad: vyhľadávanie dokumentov, vyhľadávanie na webe, filtrovanie spamu, modelovanie tém atď. Na zvládnutie takejto úlohy však nie je veľa optimálnych spôsobov. Veľa problémov používa tzv. **bag of words (BOW)** prístup, no tieto výsledky sú výsledky sú väčšinou len priemerné. Ďalšou z metód, ktorá je často používanou je Latentná Dirichletova Alokácia (LDA) slúžiaca na modelovanie (extrahovanie tém/kľúčových slov z textov), ale je veľmi ťažké ju vyladiť a výsledky sa ťažko hodnotia [9].

My sa ale pozrieme na prístup Doc2vec – NLP nástroj slúžiaci na reprezentáciu dokumentov ako vektorov a je zovšeobecnením metódy word2vec. Táto metóda je ľahká na použitie a dáva dobré výsledky. [Mikilov, Le, 2014]

4.2.1. Word2vec

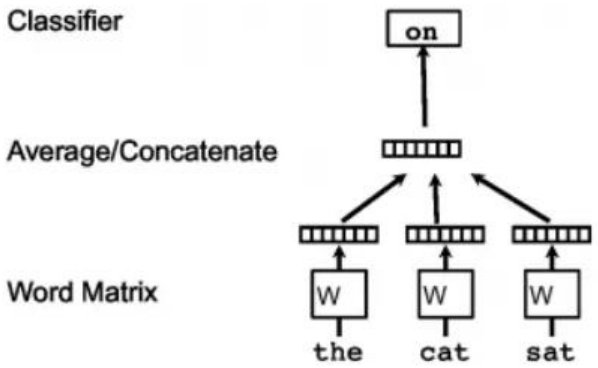
Pred tým ako začneme však potrebujeme poznať princíp word2vec. Ako už názov napovedá, word2vec je dobre známy koncept, ktorý sa používa na generovanie reprezentačných vektorov zo slov.

Vo všeobecnosti, ak chceme vytvoriť nejaký model pomocou slov, jednoducho ich označiť/rýchlo ich zakódovať pomocou nejakého id je prijateľný spôsob.

Pri použití takéhoto kódovania však slová strácajú svoj význam. napr. ak zakódujeme Paríž ako id_4, Francúzsko ako id_6 a energiu ako id_8, Francúzsko bude mať rovnaký vzťah s energiou ako s Parížom. Uprednostnili by sme reprezentáciu, v ktorej si budú Francúzsko a Paríž bližšie ako Francúzsko a energia. Word2vec nám teda dáva numerickú reprezentáciu každého slova s dôrazom na to, aby sa medzi danými slovami zachovali ich relácie.

Teda ako na to ? Reprezentácia word2vec je vytvorená pomocou 2 algoritmov: Continuous Bag-of-Words model (CBOW) a Skip-Gram model.

CBOW vytvára posuvné okno okolo aktuálneho slova, na jeho predpovedanie z „kontextu“ – okolitých slov. Každé slovo je reprezentované ako znakový vektor. Po tréningu sa tieto vektory stanú vektormi slov. Na obrázku č. 6 je možné vidieť ako CBOW používa slová ako „the“, „cat“ a „sat“ na predikciu slova „on“:

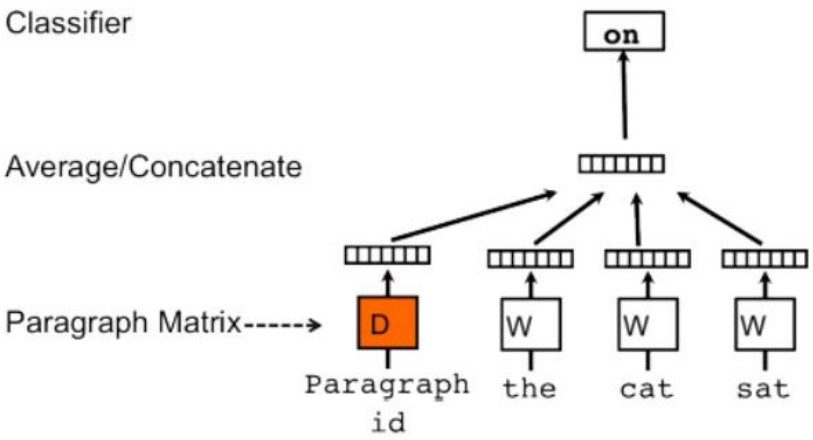


Obr. 12 CBOW

Skip gram je v skutočnosti opakom CBOW: namiesto toho, aby sme zakaždým predpovedali jedno slovo, používame 1 slovo na predpovedanie všetkých okolitých slov („kontext“). Skip gram je oveľa pomalší ako CBOW, ale považuje sa za presnejší pri zriedkavých slovách.

4.2.2. Doc2vec

Ako už bolo povedané, cieľom doc2vec je vytvoriť číselnú reprezentáciu dokumentu bez ohľadu na jeho dĺžku. Ale na rozdiel od slov, dokumenty neprichádzajú v logických štruktúrach, ako sú slová, takže je potrebné nájsť inú metódu. Koncept, ktorý použili Mikilov a Le, bol jednoduchý, ale šikovný: použili model word2vec a pridali ďalší vektor (Paragraph id) takto:

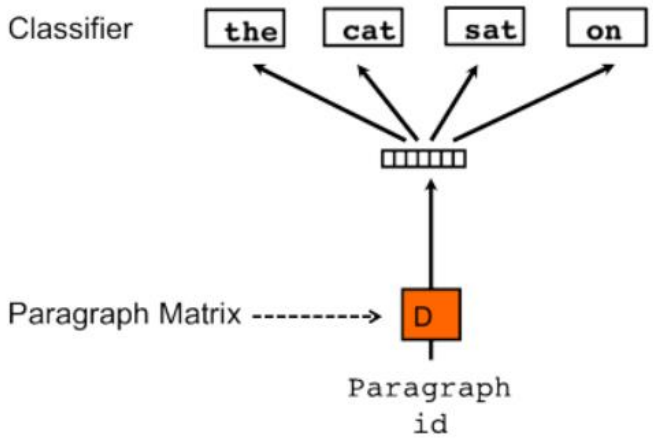


Obr. 13 PV-DM model

Ide o malé rozšírenie modelu CBOW. Ale namiesto toho, aby sme na predpovedanie ďalšieho slova používali iba slová, pridali sme aj ďalší vektor funkcií, ktorý je jedinečný pre dokumenty. Takže pri trénovaní slovných vektorov W sa trénuje aj vektor dokumentu D a na konci trénovania drží číselnú reprezentáciu dokumentu.

Vyššie uvedený model sa nazýva **verzia bodového vektora s distribuovanou pamäťou (PV-DM)**. Funguje ako pamäť, ktorá si pamätá to, čo v aktuálnom kontexte chýba. Zatiaľ čo slovné vektory predstavujú pojem slova, vektor dokumentu má za cieľ reprezentovať koncept dokumentu.

Rovnako ako vo word2vec možno použiť aj iný algoritmus, ktorý je podobný skip-gramu, **Distributed Bag of Words version of Paragraph Vector (PV-DBOW)**:



Obr. 14 PV-DBOW

Tu je tento algoritmus skutočne rýchlejší (na rozdiel od word2vec) a spotrebuje menej pamäte, pretože nie je potrebné ukladať vektory slov. Mikilov a Le odporúčajú použiť kombináciu oboch algoritmov, hoci model PV-DM je lepší a zvyčajne sám o sebe dosiahne najmodernejšie výsledky.

4.2.3. Postup implementácie doc2vec

V našom prípade je použitie prístupu word2vec nezmyselné, keďže je pre nás dôležité najskôr rozoznať, či sa jedná o škodlivú emailovú správu alebo nie. Preto potrebujeme poznať kontext textu a teda použitie doc2vec je rozumným prvým krokom.

V prvom rade potrebujeme dáta vyčistiť a normalizovať. Na nahratie dát a čistenie vieme použiť nasledujúce kódy [10]:

```
#import the libraries
import pandas as pd
df=pd.read_csv('MailDataSet.csv)
#drop the Nan rows
df.dropna(inplace=True)
```

V datasete je použitých 20 scam emailových správ a 30 legitímnych správ, všetky v anglickom jazyku. Osobné údaje prijímateľov boli patrične anonymizované. Jednotlivé emailové správy sa v dokumente vyskytujú po riadkoch pre jednoduchšiu manipuláciu a ich premenu na vektory. Nasledujúci kód následne vykoná čistenie textu [10]:

```
#import the libraries
import re
from nltk.corpus import stopwords
from nltk.stem.wordnet import WordNetLemmatizer
import gensim

lemma = WordNetLemmatizer()
stopword_set =
set(stopwords.words('english')+['a','of','at','s','for','share'
,'stock'])def process(string):
    string=' '+string+' '
    string=' '.join([word if word not in stopword_set else ''
```

```
for word in string.split()])
    string=re.sub('\@\w*',' ',string)
    string=re.sub('\.',' ',string)
    string=re.sub("[,#'-\(\):$;\?%]",' ',string)
    string=re.sub("\d",' ',string)
    string=string.lower()
    string=re.sub("nyse",' ',string)
    string=re.sub("inc",' ',string)
    string=re.sub(r'[^\x00-\x7F]+',' ', string)
    string=re.sub(' for ',' ', string)
    string=re.sub(' s ',' ', string)
    string=re.sub(' the ',' ', string)
    string=re.sub(' a ',' ', string)
    string=re.sub(' with ',' ', string)
    string=re.sub(' is ',' ', string)
    string=re.sub(' at ',' ', string)
    string=re.sub(' to ',' ', string)
    string=re.sub(' by ',' ', string)
    string=re.sub(' & ',' ', string)
    string=re.sub(' of ',' ', string)
    string=re.sub(' are ',' ', string)
    string=re.sub(' co ',' ', string)
    string=re.sub(' stock ',' ', string)
    string=re.sub(' share ',' ', string)
    string=re.sub(' stake ',' ', string)
    string=re.sub(' corporation ',' ', string)
    string=" ".join(lemma.lemmatize(word) for word in
string.split())

    string=re.sub('( [\w]{1,2} )',' ', string)
    string=re.sub("\s+",' ',string)

    return string.split()#drop the duplicate values of news

df.drop_duplicates(subset=['raw.title'],keep='last',inplace=Tru
e)

#reindex the data frame
df.index=range(0,len(df))#apply the process function to the
news titles

df['title_l']=df['raw.title'].apply(process)
df_new=df
```

Teraz, keď máme čisté údaje, môžeme tieto údaje transformovať na vektory. Implementácia knižnice gensim pre doc2vec potrebuje objekty triedy TaggedDocuments gensimu:

```
#import the modules
from gensim.models.doc2vec import Doc2Vec,
TaggedDocumentdocuments = [TaggedDocument(doc, [i]) for i, doc
in enumerate(list(df_new['title_l']))]
```

Môžeme vytvoriť náš doc2vec model:

```
model = Doc2Vec(documents, size=25, window=2, min_count=1,
workers=4)
```

Teraz máme plne načítaný model doc2vec všetkých vektorov dokumentu, ktoré sme mali v našom dátovom rámci.

Ak chceme vytlačiť všetky vektory, vieme použiť nasledujúci úsek kódu:

```
#appending all the vectors in a list for training
X=[]
for i in range(40):
X.append(model.docvecs[i])
print mdoel.docvecs[i]
```

Tieto vektory teraz obsahujú vsadenia dokumentov a sémantický význam dokumentov. Na nájdenie podobných emailových správ môžeme použiť metódy v modeli.

```
#to create a new vector
vector = model.infer_vector(process("Merger mails with
verizon"))

# to find the similarity with vector
model.similar_by_vector(vector)

# to find the most similar word to words in 2 document
model.wv.most_similar(documents[1][0])

#find similar documents to document 1
model.docvecs.most_similar(1)
```

4.2.4. Metóda zhlukovania dokumentov

Použijeme vektory vytvorené v predchádzajúcej časti na generovanie zhlukov pomocou algoritmu klastrovania K-means. Implementácia K-means je dostupná v knižnici sklearn, takže túto implementáciu vieme použiť.

```
#import the modules
from sklearn.cluster import KMeans
import numpy as np
#create the kmeans object withe vectors created previously
kmeans = KMeans(n_clusters=2, random_state=0).fit(X)

#print all the labels
print kmeans.labels_

#craete a dictionary to get cluster data
clusters={0:[],1:[],2:[],3:[]}
for i in range(40):
    clusters[kmeans.labels_[i]].append('
'.join(df_new.ix[i,'title_l']))
print clusters
```

Touto metódou vieme následne vektorizované emailové správy rozdeliť do dvoch clustrov podľa toho, či ide o podvodnú správu alebo nie.

Táto metóda je trénovateľná a použiteľná pre účely nájdenia podobností v dokumentoch a ich následnej klasifikácie. Rozdelenie podvodných správ do kategórií podľa napríklad najčastejšie sa vyskytujúcich štyroch cieľov útočníkov

- Dedičstvo
- Zoznamka
- Predvolanie na políciu
- Výhra

by bolo veľkým prínosom pre nášho konverzačného robota, keďže poznaním kategórie mailu sa hľadanie správnej odpovede stáva o to jednoduchším.

4.3. Získavanie dát pomocou manuálnych metód alebo použitím scrapovania

Scraperovanie dát (alebo web scraping) je proces automatizovaného získavania informácií z webových stránok pomocou programovacieho kódu alebo softvéru. Tento proces zahŕňa extrahovanie dát z webových stránok a ich ukladanie do štrukturovaného formátu pre ďalšie spracovanie alebo analýzu [11].

V našom prípade je web scraping jedným z veľmi rýchlych a jednoduchých spôsobov ako zrýchliť a zautomatizovať získavanie emailových správ pre trénovací dataset. Existuje množstvo webových stránok a blogov obsahujúcich množstvo podvodných správ, na ktorých vieme program trénovať.

Existuje niekoľko spôsobov, ako sa dá scrapovanie dát vykonávať. Jedným zo spôsobov je použitie programovacieho jazyka Python v kombinácii s jeho knižnicami, ako sú napríklad Beautiful Soup alebo Scrapy. Tieto knižnice umožňujú nájdenie, získavanie a spracovanie dát z webových stránok. Sú schopné extrahovať informácie z HTML alebo XML kódu stránok, ktoré môžu obsahovať text, obrázky, videá a iné prvky.

Veľkou výhodou je, že vyššie spomenuté knižnice sú voľne dostupné a nevzťahujú sa na nich žiadne autorské práva. Používateľ má plnú kontrolu nad všetkými ich funkcionalitami a metódami. Nevýhodou však môže byť nutnosť znalosti programovacieho jazyka a istá úroveň technickej zručnosti.

Ďalším spôsobom je použitie špecializovaných softvérových nástrojov, ktoré sú určené na scrapovanie dát z webových stránok. Tieto nástroje sú zvyčajne jednoduché a intuitívne, čo umožňuje aj používateľom bez rozsiahlych technických znalostí úspešne používať tieto nástroje a získavať údaje z webových stránok. Tieto nástroje zároveň poskytujú možnosť konfigurácie rôznych parametrov, ako sú napríklad filtrovanie informácií, ktoré majú byť získané, alebo zadávanie kľúčových slov na vyhľadávanie informácií.

Pre účely tejto práce nám však postačuje aj pomerne skromný dataset o veľkosti 50 emailových správ, ktoré sme zozbierali manuálne z rôznych zdrojov.

5. Záver

V nasledujúcich mesiacoch sa bližšie pozrieme na implementáciu vstupného modulu nášho konverzačného robota. Pomocou vyššie uvedených metód a nástrojov sa v prvom rade pokúsime vytvoriť funkčný model, ktorý by bol schopný identifikácie a klasifikácie základných skupín entít, obdobne ako v spomínaných nástrojoch. Následne by sme tento model rozšírili o entity, ktoré by boli prínosné pre získanie kontextu z datasetu podvodných správ, ktoré sme obdržali za posledné obdobie.

Pomocou metódy Doc2vec sa ďalej na tejto vzorke správ pokúsime natrénovať nášho chatbota tak, aby bol schopný identifikovať účel útočníka a tým dané správy klasifikoval do spomínaných štyroch kategórií, ktoré sme si zvolili.

Tento základný modul by sme následne vedeli využiť ako pomoc pri filtrovaní podvodných správ na univerzite, alebo pokračovať v jeho rozširovaní o ďalšie moduly.

6. Literatúra

1. [Online] https://www.ftc.gov/enforcement/consumer-sentinel-network/reports
2. [Online] https://www.scamwatch.gov.au/scam-statistics
3. [Online] Gaurav, G., 2018. Understanding and Leveraging Machine Learning in Conversational AI (using Rasa). Dostupné na: https://medium.com/@gauravgpunjabi/understanding-and-leveraging-machine-learning-in-conversational-ai-using-rasa-73d3136e4d16
4. Žiak, E., 2022. Rozpoznávanie entít v oblasti konverzačných robotov
5. Pool party extractor (aj pre medicínske texty): https://ner-demo.poolparty.biz/
6. [Online] Bird, Steven, Edward Loper and Ewan Klein (2009), *Natural Language Processing with Python*. O'Reilly Media Inc. https://www.nltk.org/
7. [Online] https://www.mygreatlearning.com/blog/tokenization/
8. [Online] https://spacy.io/usage/linguistic-features#named-entities
9. [Online] https://medium.com/wisio/a-gentle-introduction-to-doc2vec-db3e8c0cce5e
10. [Online] https://medium.com/analytics-vidhya/document-vectorization-301b06a041
11. Bača, M. 2023. Klasifikácia Webových stránok pomocou metód strojového učenia
12. [Online] https://www.freshworks.com/live-chat-software/chatbots/three-types-of-chatbots/
13. [Online] https://www.botreetechnologies.com/blog/wp-content/uploads/2020/12/History-of-chatbots-768x335-1.jpg
14. [Online] https://en.wikipedia.org/wiki/ELIZA#/media/File:ELIZA_conversation.png
15. [Online] https://onlim.com/wp-content/uploads/dr-sbaitso-01.jpg