TECHNICKÁ UNIVERZITA V KOŠICIACH
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Detekcia falošných správ na platformách sociálneho webu

Diplomová práca

2019 **Gabriela Demková, Bc.**

TECHNICKÁ UNIVERZITA V KOŠICIACH
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Detekcia falošných správ na platformách sociálneho webu

Diplomová práca

Študijný program:	Hospodárska informatika (HI_Ing_D)
Študijný odbor:	Hospodárska informatika (9.2.10)
Školiace pracovisko:	Katedra kybernetiky a umelej inteligencie (KKUI)
Školiteľ:	doc. Ing. Kristína Machová, PhD.

2019 Košice **Gabriela Demková, Bc.**

Abstrakt v SJ

Predložená diplomová práca prináša teoretický prehľad momentálne sa šíriaceho problému sociálnych sietí, a to falošných správ a trolingu. V tejto časti sú uvedené tipy ako s týmto problémom bojovať v prípade ich prítomnosti. Následne sú opísané konkrétne prípady falošných správ, ktoré sa objavovali už od staroveku až po súčasnosť. Hlavným ťažiskom práce je nájdenie vhodného prístupu, ktorý vie pomocou zvolenej metódy strojového učenia najpresnejšie detekovať falošné správy. Pomocou metodológie CRISP-DM je rozpracovaná praktická časť diplomovej práce, kde najdôležitejšími časťami je pri textových dátach fáza prípravy dát a modelovanie, kde sa nachádzajú experimenty, ktoré sú v závere vyhodnotené.

Kľúčové slová v SJ

Falošné správy, troling, klasifikácia, CRISP-DM, Naive Bayes, rozhodovacie stromy, náhodný les, podporné vektory

Abstrakt v AJ

The presented diploma thesis brings a theoretical overview of the currently spreading problem of social networks, namely fake news and trolling. This section provides tips how to fight this problem in the case of their presence. Subsequently, specific cases of fake news are described, that have appeared since ancient times to the present. The main focus of this work is finding a suitable approach, that is able by means of selected method of machine learning the most accurately detect fake news. The CRISP-DM methodology is used in the practical part of this diploma thesis, where the most important parts are the data preparation and modeling phase, which are evaluated at the end.

Kľúčové slová v AJ

Fake news, trolling, classification, CRISP-DM, Naive Bayes, Decision Tree, Random Forest, Support Vector Machine

TECHNICKÁ UNIVERZITA V KOŠICIACH

FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Katedra kybernetiky a umelej inteligencie

ZADANIE DIPLOMOVEJ PRÁCE

Študijný odbor: **Hospodárska informatika**

Študijný program: **Hospodárska informatika**

Názov práce:

Detekcia falošných správ na platformách sociálneho webu

Detection of fake-news in platforms of the social web

Študent: **Bc. Gabriela Demková**

Školiteľ: **doc. Ing. Kristína Machová, PhD.**

Školiace pracovisko: **Katedra kybernetiky a umelej inteligencie**

Konzultant práce:

Pracovisko konzultanta:

Pokyny na vypracovanie diplomovej práce:

1. Podať prehľad problematiky detekcie toxických príspevkov v online diskusných fórach, ako sú falošné správy a trolie príspevky.
2. Navrhnúť prístup k automatickej detekcii falošných správ založený na metódach strojového učenia.
3. Navrhnutý prístup implementovať a otestovať.
4. Vyhodnotiť úspešnosť dosiahnutých výsledkov.
5. Vypracovať dokumentáciu (hlavná časť 50 - 70 strán, prílohy - používateľská a systémová príručka, DVD s textami a softvérovými výstupmi, tlačená forma v nerozoberateľnej väzbe).

Jazyk, v ktorom sa práca vypracuje: slovenský

Termín pre odovzdanie práce: 26.04.2019

Dátum zadania diplomovej práce: 31.10.2018

prof. Ing. Peter Sinčák, CSc.
vedúci garantujúceho pracoviska

prof. Ing. Liberios Vokorokos, PhD.
dekan fakulty

Čestné vyhlásenie

Vyhlasujem, že som celú diplomovú prácu vypracovala samostatne s použitím uvedenej odbornej literatúry.

Košice, 24. apríla 2019 .....

vlastnoručný podpis

Poďakovanie

Touto cestou by som chcela vysloviť poďakovanie mojej vedúcej diplomovej práce, doc. Ing. Kristíne Machovej, PhD., za odborné vedenie, metodické usmernenie, za cenné rady a informácie a v neposlednom rade za ústretovosť a ochotu pri konzultovaní diplomovej práce. Obrovské poďakovanie patrí mojej rodine a snúbencovi za ich podporu počas celého štúdia, za povzbudenie a lásku, ktorou ma po celý čas obdarovávali.

Obsah

Zoznam obrázkov

Zoznam tabuliek

Zoznam symbolov a skratiek

USA	The United States of America
CRISP-DM	Cross Industry Standard Process for Data Mining
DTM	Matica term-dokument (Document Term Matrix)
TF-IDF	Dokumentová frekvencia - inverzná dokumentová frekvencia (Term Frequency - Inverse Document Frequency)
NB	Naive Bayes
DT	Rozhodovací strom (Decision Tree)
RF	Náhodný les (Random Forest)
SVM	Podporné vektory (Support Vector Machine)

Úvod

V dnešnej dobe technológií má k internetu prístup takmer tri a pol miliarda ľudí. Pri jeho zrode mala slúžiť pre dobro, a to s cieľom šíriť poznatky a vzdelanie, najskôr medzi akademikmi a neskôr medzi širokou verejnosťou. Keď sa neskôr začali objavovať sociálne siete ich cieľ bol podobný. Postupom času a ich rýchlym vývojom sa stali nielen komunikačným kanálom ale prostriedkom pre zdieľanie fotografií, videí, článkov, názorov a to už aj prostredníctvom mobilného telefónu. Mnoho ľudí robí veci vo svojom živote len z dôvodu, aby to mohli zdieľať na sociálnych sieťach. Bohužiaľ tento komunikačný prostriedok má aj odvrátenú stránku. Stal sa domovom falošných správ, klebiet, či nezmyslov, ktoré bohužiaľ používatelia ďalej zdieľajú bez overenia si správnosti. Každý deň je faloš a klamstvo šírené prostredníctvom sociálnych sietí z rôznych dôvodov, finančný zisk, získanie si priazne od najväčšieho počtu ľudí. A práve my, ľudia, tomu len dopomáhame.

Hlavným zámerom tejto diplomovej práce je nájsť, čo najpresnejší algoritmus strojového učenia, ktorý by dokázal detekovať práve tieto falošné správy.

Práca je rozdelená na štyri kapitoly a to, formulácia úlohy, teoretická časť, analýza súčasného stavu a praktická časť.

Prvá kapitola - *Formulácia úlohy a cieľ práce* - stručne opisuje zadanie tejto diplomovej práce.

V druhej kapitole - *Teoretický rozbor zvolenej témy* - sa nachádza teoretické vymedzenie sociálneho webu, jeho história a opis najrozšírenejších sociálnych sietí v dnešnej dobe. Na záver sú opísané jeho výhody a nevýhody. Ďalšou časťou je popis trolingu, typov trolov a boj proti nim. Treťou časťou teoretického vymedzenia je opis falošných správ, ich šírenie od stredoveku až po súčasnosť.

Tretia kapitola - *Analýza súčasného stavu* - obsahuje rozpracovanie viacerých prípadových štúdií, ktoré sa zaoberali detekciou falošných správ.

Štvrtá kapitola - *Praktická časť* - sa venovala riešeniu úlohy danej diplomovej práce, kde sme pracovali s algoritmami strojového učenia, ktoré sme aplikovali na dve dátové množiny, ktoré obsahovali falošné aj pravdivé správy. Pri riešení sme postupovali pomocou metodológie CRISP-DM a pracovali sme v prostredí R studio s programovacím jazykom R. Dáta sme v prvom rade upravili a následne sme vytvorili modely, ktoré sme na záver vyhodnotili pri sledovaní viacerých kritérií.

1. Formulácia úlohy a cieľ práce

Úlohou študenta pri vypracovaní tejto diplomovej práce je:

1) **Podať prehľad problematiky detekcie toxických príspevkov v online diskusných fórach, ako sú falošné správy a trolie príspevky.**

   Tento bod v sebe zahŕňa teoretický prehľad sociálneho webu, jeho vývoj, výhody a nevýhody. Ďalej sme opísali troling, typy trolov a rady ako proti nim bojovať. Teoreticky sme opísali falošné správy a ich vplyv od minulosti až po súčasnosť. V neposlednom rade sme vytvorili analýzu súčasného stavu detekcie falošných správ.

2) **Navrhnúť prístup k automatickej detekcii falošných správ založený na metódach strojového učenia.**

   Táto časť zahŕňa výber vhodných algoritmov strojového učenia, ktoré budú slúžiť pre najlepšiu detekciu falošných správ.

3) **Navrhnutý prístup implementovať a otestovať.**

   V tejto časti začína praktická časť diplomovej práce, v ktorej sme vybrané metódy spracovali a ich funkčnosť otestovali.

4) **Vyhodnotiť úspešnosť dosiahnutých výsledkov.**

   Bod, v ktorom sme dosiahnuté výsledky vyhodnotili pomocou viacerých sledovaných kritérií.

5) **Vypracovať dokumentáciu (hlavná časť 50 - 70 strán, prílohy - používateľská a systémová príručka, DVD s textami a softvérovými výstupmi, tlačená forma v nerozoberateľnej väzbe).**

2. Teoretický rozbor zvolenej témy

Teoretický rozbor zvolenej témy pozostáva z rozboru troch okruhov, ktorými sú sociálny web, troling a falošné správy. V časti sociálneho webu je okrem definície opísaná aj jeho história, príklady najpopulárnejších sociálnych médií, jeho vývoj a taktiež jeho výhody a nevýhody. Druhá časť sa zaoberá opisom trolingu, ako aj miestami, kde sa troling najčastejšie objavuje. Okrem toho v tejto časti opíšeme typy trolov, s ktorými sa na sociálnych médiách stretávame a uvedieme pár rád ako proti nim bojovať, ak sme sa už v ich prítomnosti ocitli. V poslednej časti teoretického rozboru definujeme pojem falošných správ taktiež spolu s ich typmi. Časovo uvedieme pár zaujímavostí, ktoré boli spôsobené šírením falošným správ od staroveku po súčasnosť.

2.1. Sociálny web

Sociálny web je definovaný ako súbor sociálnych vzťahov, ktoré spájajú ľudí prostredníctvom World Wide Web, laicky povedané prostredníctvom internetu. Taktiež zahŕňa webové stránky, softvér, hardvér a systémy, ktoré sú navrhnuté, vytvorené a spustené na podporu tejto sociálnej interakcie. Tieto sociálne interakcie tvoria základ mnohých online aktivít vrátane nakupovania, vzdelávania, hrania hier, zdieľania hudby, výmeny správ a informácií. Sociálne siete ako Facebook, Twitter, Linkedln, Pinterest, Instagram a mnoho iných sú toho súčasťou. [1]

Keďže aktivita ľudí na internete a komunikácia medzi používateľmi narastá, informácie o ich sociálnych vzťahoch sú čoraz viac dostupné. Sociálne siete v súčasnosti umožňujú ľuďom a organizáciám navzájom komunikovať. Stovky miliónov používateľov internetu využíva tisíce sociálnych webových stránok, aby ostali v kontakte so svojimi priateľmi, vytvárali si prostredníctvom nich nové priateľstvá a zdieľali s nimi svoje fotografie, videá, články, svoje názory a ukázali im, čo sa im páči a to už dokonca aj prostredníctvom podpory mobilných platforiem pre mobilné telefóny. Sociálny web presahuje jednoduché webové aplikácie, ktoré jednotlivcov spájajú s úplne novým spôsobom života. Napríklad, spoločnosť Facebook oznámila v roku 2017 až 1,86 miliardy používateľov, služba YouTube 100 miliónov videí a tieto čísla sa neustále zvyšujú. [1]

2.1.1. História sociálneho webu

Rovnako ako telefón, internet nebol vytvorený len ako nástroj na komunikáciu, ale vyvinul sa na to, aby bol súčasťou každodenného života. Rozvíjal sa v troch etapách a to od začiatku 90-tych rokov až po súčasnosť sa transformoval z jednosmerných komunikačných webových stránok do siete skutočných sociálnych aplikácií. V polovici 90-tych rokov, počas éry takzvanej „jednosmernej konverzácie“ bola väčšina webových stránok postavených výhradne tak, že informácie tiekli od osoby alebo organizácie, ktorá prevádzkovala danú stránku. Komunikácia bola veľmi ťažká a dosiahla sa

výlučne prostredníctvom jednotlivcov, ktorí na svoje príspevky odpovedali na svojej webovej stránke. V polovici 90-tych rokov zaznamenala spoločnosť Amazon zlepšenie v online sociálnej interakcii tým, že zistili ako prepojiť databázy s ich webovými stránkami s cieľom ukladať informácie a následne ich zobrazovať. V spojení s inými inováciami to umožnilo „obojsmernú konverzáciu“ medzi používateľmi a jednotlivcom alebo organizáciou. Webové stránky sa časom stali sofistikovanejšie, ľudia sa stali pohodlnejšími a rozšíril sa taktiež prístup na internet, čo smerovalo k tomu, že návrhári začali implementovať nové funkcie, ktoré umožňovali komunikovať nielen s vydavateľmi webových stránok ale aj s inými používateľmi, ktorí danú stránku navštívili. Bol to obrovský spoločenský krok, ktorý po prvýkrát umožnil interakciu v skupine. Práve vďaka tomu sa začala rozlišovať webová aplikácia od sociálnej webovej aplikácie. Prvé stránky sociálnych sietí boli vytvorené ešte pred zavedení sociálnych médií. Tvrdí sa že prechod na stránky sociálnych médií sa začal po vytvorení prvej svetovej komunite interaktívneho denníka Open Diary, ktorý bol založený v roku 1998 a využíva sa dodnes. Open Diary spojil do jednej komunity spisovateľov týchto denníkov a považovalo sa to za rannú éru stránok sociálnych sietí. V tom čase bol vytvorený termín „weblog“, ktorý bol neskôr skrátený na „blog“. Práve to sa považuje za začiatok súčasného obdobia sociálnych médií, pričom tento termín vstupuje do bežného používania, nakoľko vysokorýchlostný internet sa stáva čoraz viac dostupným a to vedie k vzrastu sociálnych sietí. [2]

Popis najpopulárnejších sociálnych médií: [1], [3]

Facebook - obľúbená bezplatná webová stránka sociálnych sietí, ktorá umožňuje registrovaným používateľom vytvárať profily, nahrávať fotografie a videá, hrať hry, posielať správy a udržiavať kontakt s priateľmi, rodinou a kolegami.

Twitter - bezplatná mikroblogovacia sociálna sieť, ktorá umožňuje svojim registrovaným používateľom posielať a čítať krátke správy ostatných používateľov nazývané tweety, pomocou viacerých platforiem a zariadení.

Wikipedia - bezplatná, online encyklopédia. Ktokoľvek, kto je registrovaný môže vytvoriť článok na publikovanie. Upravovať články môžu aj nezaregistrovaný používatelia.

LinkedIn - sociálna sieť vytvorená špeciálne pre podnikateľskú komunitu. Cieľom stránky je umožniť registrovaným členom vytvorenie profesionálnej siete. Profil používateľa obsahuje jeho vlastný životopis.

Reddit - sociálna spravodajská webová stránka s fórum, kde sú príbehy propagované členmi stránky. Táto stránka sa skladá zo stoviek podsúborov, kde každý má špecifickú tému, ako je technológia,

hudba alebo politika. Členovia predkladajú obsah a ostatní členovia hlasujú. Cieľom je najlepšie príbehy umiestniť na začiatok hlavnej stránky.

Pinterest - internetová stránka pre zdieľanie a kategorizáciu obrázkov nájdených online na internete. Pinterest vyžaduje krátky popis k obrázkom. Kliknutím naňho sa dostaneme k pôvodnému zdroju. Keď napríklad klikneme na obrázok obuvi, navedie nás to na stránku, kde si danú obuv môžeme objednať. Kliknutím na palacinky nás privedie k receptu.

2.1.2. Vývoj sociálneho webu

Sociálny web sa veľmi rýchlo stáva životným štýlom, mnoho ľudí navštevuje stránky sociálnych sietí minimálne raz do dňa. Navyše ohromne rýchly rast sociálneho webu od 90.tych rokov nepredpokladá v blízkej budúcnosti spomalenie. Čiara medzi sociálnymi sieťami a sociálnymi médiami sa stáva čoraz viac tenšou, keďže stránky sociálnych sietí obsahujú fotografie, videá a ďalšie funkcie, ktoré sú typické pre stránky sociálnych médií, rovnako ako stránky sociálnych médií integrujú charakteristické črty stránok sociálnych sietí do svojich vlastných online rámcov. Jednou významnou zmenou, ktorú prinieslo zlúčenie sociálnych médií a sietí je transformácia sociálnych webových aplikácií do egocentrického softvéru, ktorý dáva ľuďom do centra aplikácií. Moderný softvér sociálneho webu vytvára širší súbor spoločenských interakcií, ktoré sú dostupné pre používateľov, ako napríklad priateľstvo, sledovanie jednotlivcov, či dokonca posielanie virtuálnych darov alebo bozkov. Sociálne webové aplikácie sa zvyčajne vytvárajú pomocou objektovo orientovaného programovania s využitím kombinácie viacerých programovacích jazykov. [2]

Život používateľov sociálneho webu je čoraz viac prepojený s ich online profilmi a účtami, a to do takej miery, že mnohé sociálne siete a stránky sociálnych médií momentálne ponúkajú podporu pre mobilné zariadenia a pripojenie na internet. Populárne sociálne webové stránky umožnili svojim používateľom zdieľať nový obsah s ostatnými, aktualizovať svoje statusy a dostávať aktualizácie o svojich priateľov prostredníctvom mobilných platforiem. Pre používateľov je veľmi dôležité udržiavať kontakt so svojimi priateľmi online po celú dobu a aktualizovať svoj profil aj vtedy, keď sa nachádzajú mimo svojich používateľov. [1]

Webové stránky, ktoré nie sú založené na sociálnych interakciách však pridávajú funkcie, ktoré umožňujú diskusiu a spoluprácu o rozšírenie svojich užívateľských základov. Už v roku 1995 elektronický predajca Amazon realizoval takéto funkcie, najmä zákaznícke recenzie s veľkým úspechom. Zákaznícke recenzie sa správajú ako magnet a lákajú ľudí na stránku. Poskytujú cenné informácie, ktoré jednotlivci hľadajú a sú napísané používateľmi zadarmo a jednoducho s cieľom zdieľať svoje skúsenosti o produkte alebo službe s ostatnými bez akýchkoľvek potenciálne zaujatých informácií. [1]

Webové stránky s osobitným záujmom taktiež implementovali funkcie sociálneho webu na ich rozšírenie. Jedným z príkladov je komunita 10 miliónov kuchárov, ktorí si navzájom vymieňajú nápady a recepty. Okrem výmeny receptov s ostatnými prostredníctvom webovej stránky môžu používatelia hodnotiť a publikovať recenzie vyskúšaných receptov a poskytnúť návrhy na ich zmenu alebo zlepšenie. Spätná väzba sa využíva na hodnotenie a klasifikáciu receptov. Stránka taktiež rozšírila svoje služby tým, že zahrnula sociálne funkcie ako sú blogy používateľov a pripojila sa k iným sociálnym sieťam a médiám, aby rozšírila svoju prítomnosť na sociálnom webe. Recepty nájdené na tejto webovej stránke sa stávajú súčasťou sociálnej siete, nakoľko ich ostatní členovia hodnotia, komentujú a poskytujú spätnú väzbu, prečo bol daný recept dobrý alebo zlý, alebo zdieľali spôsoby akými by ich zmenili. [1]

2.1.3. Výhody a nevýhody sociálnych médií pre spoločnosť

Sociálne média sa v posledných rokoch veľmi rýchlo rozrástli. Špeciálne stránky sociálnych sietí ako Facebook a Twitter sa rozrástli o miliónov používateľov v priebehu niekoľkých rokov. Spôsob akým rastie technológia je zrejmé, že čoraz viac používateľov vidí jej výhody. Pre spoločnosť priniesli množstvo výhod. Od pokročilých štátov, až po menej rozvinuté krajiny, každý národ využíva silu sociálnych médií na zlepšenie života a taktiež ich použiť na horkosť ľudí. [3]

Avšak na druhej strane sociálne médiá negatívne ovplyvňujú spoločnosť. Rovnako ako čokoľvek iné, všetko, čo môže byť použité ako dobré, môže byť aj zlé. Taktiež aj sociálne médiá poskytli spoločnosti pozitíva aj negatíva. Je to všetko o tom, ako používame a robíme veci pozitívne prostredníctvom sily sociálnych médií. Je to v rukách používateľa, aby využil sociálne médiá vo svoj prospech. [3]

Výhody sociálnych médií: [3]

Konektivita - hlavnou výhodou sociálnych médií je rýchlosť pripojenia. Ľudia z odkiaľkoľvek sa môžu pripojiť s kýmkoľvek, bez ohľadu lokality či náboženstva. Krása sociálnych sietí spočíva v tom, že sa môžete spojiť s kýmkoľvek, aby ste sa naučili a zdieľali svoje myšlienky.

Vzdelávanie - sociálne média prinášajú množstvo výhod pre študentov a učiteľov. Je veľmi jednoduché sa vzdelávať od iných, ktorí sú odborníkmi prostredníctvom sociálnych médií.

Pomoc - môžete zdieľať svoje problémy s komunitou, aby ste získali pomoc. Či už potrebujete pomoc z hľadiska peňazí alebo psychológie, môžete ju získať z komunity, s ktorou ste spojení.

Informácie a aktualizácie - ďalšou výhodou sociálnych médií je to, že sú aktualizované o najnovšie udalosti sveta. Väčšinou televízie a tlačové médiá poskytujú skreslené alebo nepravé správy.

S pomocou sociálnych médií môžete získať fakty a pravdivé informácie vykonaním nejakého výskumu.

Propagácia - či už ste online alebo offline, môžete propagovať svoju firmu najväčšiemu publiku.

Šľachetná príčina - sociálne médiá môže byť použité aj pre charitatívne činnosti. Napríklad na podporu mimovládnych organizácií, darcovstva pre ľudí v núdzi a na aktivity v oblasti sociálnej starostlivosti. Ľudia využívajú médiá na darcovstvo pre ľudí v núdzi a môže to byť veľmi rýchly spôsob, ako im pomôcť.

Povedomie - sociálne média taktiež vytvárajú povedomie a inovujú spôsob, akým ľudia žijú. Pomáha ľuďom objaviť nové a inovatívne témy, ktoré môžu zlepšiť ich osobné životy. Od poľnohospodárov po učiteľov, študentov práva, každý jednotlivec spoločnosti môže mať prospech zo sociálnych médií a jeho faktoru povedomia.

Pomáha vládam a polícií v boji proti kriminalite - jednou z výhod sociálnych médií je taktiež fakt, že pomáha vládam a bezpečnostným agentúram špehovať, chytať zločincov a bojovať tak proti kriminalite.

Zvyšuje obchodnú reputáciu - pozitívne komentáre a zdieľanie informácií o podniku môže pomôcť s predajom a reputáciou podniku. Keďže ľudia môžu slobodne zdieľať, čo chcú na sociálnych médiách, môže to mať veľmi pozitívny vplyv pri zdieľaní dobrých informácií.

Pomáha budovať komunity - náš svet má mnoho náboženstiev a presvedčení. Sociálne médiá pomáhajú budovať a zúčastňovať sa v spoločenstve vlastného náboženstva a veria, že sa o ňom dozvedia, čo najviac prostredníctvom diskusie. Podobne sa ľudia z rôznych komunít môže pripojiť, aby diskutovali a zdieľali informácie k tomu súvisiace. Napríklad milovníci nejakej konkrétnej hry sa pripoja k spoločenstvu súvisiacou s danou hrou, je to možné pri akejkoľvek téme.

Nevýhody sociálnych médií: [3]

Kybernetické šikanovanie - v dnešnej dobe je veľmi jednoduché vytvoriť si falošný účet a robiť čokoľvek bez toho aby bol vysledovateľný. Hrozby, zastrašovacie správy a klebety môžu byť zaslané ako hromadné správy aby vytvorili nepokoj a chaos v spoločnosti.

Hacking - osobné údaje a súkromie môže byť veľmi ľahko napadnuté a zdieľané na internete. To môže spôsobiť finančné straty, či straty osobného života. Rovnako krádež identity je ďalšou záležitosťou, kde hacker zasiela materiály, ktoré trpko ovplyvňujú osobný život používateľa. Toto je jedna z nebezpečných nevýhod sociálnych médií a každému používateľovi sa odporúča udržať svoje osobné údaje a účty v bezpečí, aby sa vyhli takýmto typom nehôd.

Závislosť - návyk na sociálne médiá je veľmi zlý a môže taktiež narušiť osobný život. Čoraz viac mladých ľudí je závislých na sociálnych médiách. Veľmi intenzívne sa závislosť podieľa na odrezaní ľudí od spoločnosti. Strácame čas, ktorý by mohol byť využitý produktívnymi aktivitami.

Podvodníci a podvody - je k dispozícií niekoľko príkladov , kedy sa jednotlivci stali terčom podvodu prostredníctvom sociálnych médií. Tieto podvody sa pravidelne opakujú: skrytá adresa URL, neoprávnené získavanie údajov, skryté poplatky, zmocnenie sa cudzích peňazí, šírenie nepravdivých správ.

Bezpečnostné problémy - bezpečnostné agentúry majú prístup k osobným účtom používateľov. To spôsobuje, že súkromie je ohrozené. Nikdy neviete, kedy vás navštívi vyšetrovací úradník, pokiaľ ide o akýkoľvek problém, ktorý ste omylom alebo nevedomky diskutovali prostredníctvom internetu.

Reputácia - sociálne médiá môžu ľahko zničiť povesť niekoho iného len tým, že vytvoria falošný príbeh a rozšíria ho po sociálnych sieťach. Podobne môžu dopadnúť aj podniky v dôsledku šírenia zlej povesti.

Problémy so vzťahmi a neverou - väčšina ľudí používa platformu sociálnych médií na to, aby sa vzali. Po určitom čase si však uvedomujú nesprávne rozhodnutie. Rovnako sa páry podvádzajú tým, že zobrazujú falošné pocity a nepravdivé informácie.

Zdravotné problémy - nadmerné využívanie sociálnych médií môže mať negatívny vplyv na zdravie. Cvičenie je kľúčom k chudnutiu a zdravému životnému štýlu, no väčšina ľudí sa stáva lenivým práve kvôli nadmernému používaniu sociálnych médií. V dnešnej dobe je zdravie kvôli nim veľmi ovplyvnené.

Sociálne médiá spôsobujú smrť - nie len ich používaním ale aj sledovaním bláznivých videí, ktoré sú zdieľané na internete. Napríklad motorkári napodobňujú kaskadérske kúsky, ľudia skáču po na idúcich vlakoch. Tieto typy kaskadérskych kúskov robia mladí ľudia aby boli zaujímaví a aby sa podobali iným.

Idealizovanie drog a alkoholu - ďalšou z nevýhod sociálnych médií je fakt, že používatelia sledujú bohatých a drogovo závislých ľudí, ktorí svoje názory zdieľajú na internete. To inšpiruje ostatných, ktorý ich následne napodobňujú.

2.2. Troling

„Trols“ je internetový slang označujúci osoby, ktoré sa úmyselne pokúšajú podnecovať konflikt alebo nepriateľstvo tým, že uverejňujú urážlivé, zápalové, provokatívne alebo irelevantné komentáre v online sociálnom spoločenstve. Ich zámerom je rozrušiť ostatných a vyvolať silnú

emocionálnu odpoveď, najideálnejšie negatívnu. To využívajú ako návnadu pre zapojenie sa nových používateľov do diskusie. Avšak nie každý nahnevaný používateľ internetu je trol. Online svet je plný nadšených používateľov, ktorí chcú zdieľať svoje myšlienky. Rozdiel medzi nimi a trolmi je ten, že aj keď sa používateľ nahnevá alebo reaguje agresívne, verí tomu, čo vyjadruje. Trol nemusí veriť tomu, čo zdieľa. Vybral si to z dôvodu rozrušenia ostatných účastníkov na sociálnych sieťach. Činnosť trola je v internetových diskusiách označovaná ako troling. Bližšie môžeme konštatovať, že troling je aktivita s asociálnym jednaním využívajúca citlivosť spoločnosti na šírenie fanatizmu, rasizmu, nenávisti alebo vyvolávajú konflikty hašterením medzi inými, častokrát na veľmi provokatívne témy. Napriek pokusom o obmedzenie je troling stále viac a viac rozšírený. Môžeme sa s ním stretnúť vo všetkých zákutiach sociálneho webu, kde ľudia môžu slobodne vyjadrovať svoje názory a myšlienky. [4], [5], [6], [7]

Uvedieme konkrétne miesta, ktoré lákajú osoby aby rozširovali troling: [6]

Komentáre k videám na YouTube - YouTube je známy tým, že obsahuje najhoršie komentáre zo všetkých stránok sociálnych médií. Pri každom ľubovoľnom videu sa nájde pár poburujúcich komentárov. Čím viac komentárov a názorov video obsahuje, tým viac to láka trolov pre zverejnenie práve svojich komentárov.

Komentáre k blogu - na niektorých populárnych blogoch a spravodajských stránkach, ktoré majú povolené komentovanie článkov je možné nájsť trolov, čo následne spôsobuje problémy. Platí to najmä pre blogy, ktoré pokrývajú kontroverzné témy alebo tie, ktoré majú tendenciu vyzdvihnúť veľa komentárov od ľudí, ktorí chcú zdieľať svoje názory.

Fóra - fóra sú vytvorené za účelom diskusie o témach s rovnako zmýšľajúcimi ľuďmi, čo priťahuje trolov za účelom odpútania diskusie za pomoci negatívnych slov. Ak ich administrátori fóra nezablokujú, ostatní členovia začnú odpovedať a skôr ako si to všimnú sa diskusia dostane úplne mimo témy a stane sa len jednou nezmyselnou diskusiou.

E-mail - existuje veľa trolov, ktorí aktívne venujú svoj čas a energiu aby napísali hrozné e-mailové správy ako odpoveď na ľudí, s ktorými nesúhlasili alebo nimi boli urazení.

Facebook, Twitter, Instragram a prakticky ktorákoľvek stránka sociálnej siete - v dnešnej dobe môže ktokoľvek komentovať aktuálny status na Facebooku, odpovedať na tweet, konverzovať v komunite alebo poslať anonymnú otázku, troling je absolútne všade, kde ľudia môžu používať interakciu. Instagram je na to obzvlášť negatívny, nakoľko je to verejná platforma, ktorú ľudia využívajú na uverejňovanie fotografií, pozývajú každého na ich prezeranie a posúdenie ich vzhľadu v sekcii komentárov.

Anonymné sociálne siete - slúžia ako pozvánka pre trolov, pretože sa používatelia nemusia obávať toho, že sa ich identita bude spájať s asociálnym správaním. Môžu vyjadriť svoj hnev a nenávisť bez toho, aby za to znášali dôsledky, pretože sa môžu schovať za meno niekoho iného.

A prečo sa vlastne ľudia stávajú trolmi? Jedným zo spôsobov ako porozumieť tomu, prečo sa ľudia zapájajú do trolingu je preskúmať, či prejavia osobitné osobnostné črty, ako je narcizmus, ktorý je spojený s pocitmi nadradenosti, psychopatia spojená s impulzívnosťou a bezúhonnosťou, Machiavellianismu spojený s manipuláciou a vykorisťovaním iných alebo sadizmus, ktorý je definovaný ako ukrutnosť prejavujúca sa trýznením iných. Tieto znaky bežne spočívajú v mnohých formách sociálnej manipulácie a podvodu a zahŕňajú úsilie o bezohľadné seba projektovanie, agresiu a najmä nedostatok empatie a ťažkej nevoľnosti. V štúdii z roku 2014 sa zistilo, že ľudia s vyšším stupňom sadizmu, psychopatie a Machiavellianismu majú väčšiu pravdepodobnosť byť zapojení do online trolingu, pričom sadizmus je najsilnejšou črtou. [4]

V ďalšom výskume sa zistilo, že jednotlivcom, ktorí sa zapojili do online trolingu ide o odmenu, konkrétne o sociálne uznanie. Existujú dve formy sociálnych odmien a to typické a atypické. Typické sociálne odmeny sa vo všeobecnosti vyskytujú prostredníctvom vzájomného pozitívneho spoločenského uznania a interakcií. Atypické sociálne odmeny známe aj ako negatívna spoločenská sila sa meria pomocou dotazníka o sociálnych odmenách, v ktorom účastníci naznačujú, že súhlasia s tvrdeniami ako napríklad „Mám rád, keď sa niekto nahneval“ alebo „Mám radosť z trápenia ostatných“. Práve to je pre nich pozitívnym pocitom, ktorí niektorí ľudia zažívajú pri vytváraní spoločenského nesúladu pomocou sebeckého správania a interakcií. Jednotlivci, ktorí hľadajú negatívnu spoločenskú silu si budú pravdepodobne užívať utrpenie a bolesť ostatných. V tomto výskume zhromaždili vzorku 396 dospelých, z toho približne 76% žien a 24% mužov a požiadali ich aby vyplnili dotazník na meranie úrovne narcizmu, psychopatie, Machiavellianismu a sadizmu. Posúdili taktiež aj ich orientáciu na negatívnu spoločenskú silu a ich zapojenie sa do trolingu na Facebooku. Výsledky výskumu ukázali, že najsilnejším faktorom je sadizmus a taktiež, že muži sa zapojili do trolingu na Facebooku viac ako ženy. Prekvapujúce však bolo, že výsledok taktiež ukázal, že vplyv negatívnej spoločenskej sily bol výrazne vyšší ako účinky sadizmu a psychopatie. To znamená, že zatiaľ čo osobnostné črty zohrávajú určitú úlohu, to, čo skutočne ovplyvňuje správanie trolov je sociálne potešenie z toho, že ostatní účastníci sú nahnevaní. Čím viac sociálneho potešenia trol má, tým viac sa ich správanie posilňuje. Z tohto výskumu vyplýva, že ak trolovia nedostanú negatívnu sociálnu odmenu, potom sa ich motivácia zapojiť sa ďalej do diskusie zmenšuje.

Iným dôvodom prečo sa používatelia zapájajú do trolingu je anonymita. Väčšina online sociálnych webov umožňujú používateľom vytvárať mená, ktoré ani nie sú spojené s ich totožnosťou.

Niektorí práve využívajú túto anonymitu z dôvodu, že nebudú kvôli ich správaniu nijako postihnutí. [5]

Ďalším z dôvodov je desenzibilizácia, ktorá spôsobuje znecitlivenie na veci, ktoré vidíme alebo počujeme. To je zapríčinené trávením množstva času v prostredí, ktoré nás ovplyvňuje. Zvykneme si na ne a už si viac nezasluhujú našu pozornosť. Vidieť kruté a šokujúce komentáre v online svete sa môžu postupom času zdať normálne. Používateľ, ktorý kedysi premýšľal o dôsledku svojich činov, môže teraz zaslať necitlivý alebo nevhodný komentár bez toho, aby si uvedomil, že urobil niečo zlé. [5]

2.2.1. Typy trolov

Bez ohľadu na to, kde nájdeme trolov, ktorí číhajú, všetci majú tendenciu narušovať komunity veľmi podobným a často predvídateľným spôsobom. Opíšeme si niektoré typy najčastejšie sa vyskytujúcich trolov, s ktorými sa stretávame v aktívnych online komunitách: [8]

Trol, ktorý uráža

Trol, ktorý uráža ostatných je čistý nenávisti, prostý a jednoduchý a to ani nemusí mať dôvod nenávidieť alebo urážať iných používateľov. Tento typ trolu si často ako svoju obeť vyberá každého, oslovuje ho menom, obviňujú ich z nejakých vecí a robia všetko preto, aby od nich získali negatívnu emocionálnu odpoveď. V mnohých prípadoch môže byť tento druh trolingu tak závažný, že môže viesť alebo byť považovaný za vážnu formu kyberšikany.

Trol, ktorý vytrvá v diskusii

Tento typ trolov miluje dobré argumenty. Môžeme mať skvelý, dôkladne preskúmaný a faktom založený obsah, trolovia prichádza zo všetkých protikladných uhlov do diskusie aby napadli jeho posolstvo. Veria, že majú pravdu iba oni a všetci ostatní sa mýlia. Často je ich ľahko spozorovať tým, že opustia dlhé vlákna komentárov a sú vždy rozhodnutí mať posledné slovo, pokračujúc v komentári, kým sa iný používateľ nevzdá.

Trol, ktorý kontroluje pravopis a gramatiku

Sú to ľudia, ktorí musia vždy iným používateľom povedať, že majú nesprávne napísané slová alebo gramatické chyby. Dokonca ja keď to robia jednoduchým komentovaním s opraveným slovom za symbolom hviezdičky, je to nevítaný komentár v akejkoľvek diskusii. Niektorí z nich dokonca požívajú chyby gramatiky a pravopisu ako zámienku pre urážku ostatných.

Trol, ktorý je stále urazený

Keď sú kontroverzné témy diskutované online, častokrát to smeruje k tomu, že niekto niekoho urazí, to je normálne. Ale existuje typ trolov, ktorí vypustia do diskusie vtip, paródiu alebo niečo sarkastické. Sú odborníkmi na to, aby vzali humorný obsah a premenili ich na tvrdenia voči svojej osobe, čím sa stávajú obeťou.

Trol, ktorí sa chváli a všetko vie

Tento typ trolu sa nemusí nutne zúčastňovať argumentov ale zbožňuje zdieľať o sebe extrémne detaily a fámy, dokonca šíriť svoje tajomstvá. Milujú počúvať samy seba a sú neskutočne radi, keď sa môžu zúčastniť diskusie, kde píšu len o tom, čo vedia a je im jedno, či to niekto číta alebo nie.

Trol, ktorý používa Caps Lock

Trol, ktorý používa Caps Lock nemá žiaden argument aby sa pripojil do diskusie a využíva len tlačidlo veľkých písmen. V mnohých prípadoch je tento typ trolu len nudné dieťa, ktoré hľadá niečo, čo môže robiť bez toho aby na to vynaložil nejaké úsilie a snahu. Na druhej strane obrazovky je niekto naozaj neškodný.

Trol, ktorý používa len jedno slovo

V online svete sa nájde vždy jeden prispievateľ, ktorý buď na Facebooku alebo Instagrame okomentuje príspevok jedným slovom, a to napríklad slovami ako „lol“, „áno“, „čo“ alebo „nie“. Určite nie sú najhorším typom trolov, ktorých poznáme, ale keď sa diskutuje o vážnej alebo podrobnej téme, ich jednoslovné odpovede sa stávajú len nepríjemnosťou pre tých, ktorí sa snažia sledovať diskusiu alebo pridať nejakú hodnotu do diskusie.

Trol, ktorý preháňa

Trol, ktorý preháňa môže byť niekedy kombináciou trola, ktorý všetko vie, trolom urazeným, či dokonca trolom, ktorý vytrváva v diskusii. Vedia, ako uchytiť problém, či debatu a prispôsobiť si ju. Niektorí z nich sa pokúšajú zabaviť, zatiaľ čo iní sa snažia debatu znepríjemniť. Málokedy prispievajú skutočnou hodnotou do diskusie, práve naopak, vyvolávajú problémy, ktoré absolútne s danou debatou nesúvisia.

Trol, ktorý je mimo témy

Je dosť ťažké si obľúbiť niekoho, kto zverejňuje niečo úplne mimo témy v akejkoľvek spoločenskej diskusii. Je to ešte horšie, keď sa mu podarí zmeniť tému a všetci sa ňu začnú reagovať a diskutovať

o niečom úplne inom. Dá sa to vidieť momentálne všade v online svete - v komentároch na Faceboku, Instagrame, či pri videách na YouTube, doslova kdekoľvek, kde sa uskutočňujú online diskusie.

Trol, ktorý je nenásytným spammerom

V neposlednom rade opíšeme trola, ktorý neustále spammuje ostatných používateľov online diskusií. Tento typ trolu sa nezaujíma o vaše príspevky či diskusiu, prispieva jedine pre svoj prospech. Chce aby ste navštívili jeho stránku, kúpili niečo z jeho odkazu, či využili jeho kupón. Títo trolovia taktiež rozbíjajú diskusie na sociálnych stránkach príspevkami „sleduj ma".

2.2.2. Rady a tipy pre riešenie trolov

Online zneužívanie je sociálny problém, ktorý sa jednoducho deje z dôvodu stále sa rozširujúcich technológií. Tak ako trolovia vedia ublížiť jednotlivcom, čoraz viac spôsobujú problémy aj veľkým či malým spoločnostiam. Pre nich sa sociálne média stali novým spôsobom ako získať nových zákazníkov, poskytnúť im podporu, uviesť na trh nové produkty, zdieľať správy a dokonca rozvíjať vzťahy so svojimi zákazníkmi. Vzhľadom na voľný prístup sociálnych sietí, môžu veľké spoločnosti využívať výhody osobných interakcií. Spoločnosť môže označiť „páči sa mi to" príspevok svojho spokojného zákazníka. Naopak zákazník sa môže sťažovať ak je s produktom nespokojný. Tu sa objavuje priestor pre trolov, ktorí chcú opakovanými komentármi poškodiť dobré meno spoločnosti. Uvedieme tipy a rady ako riešiť prítomnosť trolov: [9]

Nepoužívať automatické odpovede na opodstatnené sťažnosti

Jedna z najhorších vecí, ktorú môže urobiť každá spoločnosť pri riešení sťažností na sociálnych médiách a komentárov od trolov je zaslať automatickú odpoveď. Čím viac spoločnosť rastie, tým viac je potrebné vynakladať čoraz väčšie úsilie o skutočnú interakciu so zákazníkmi. [9]

Vyhradiť si čas na vypracovanie odpovedí na skutočné problémy

Najideálnejšie je vyriešiť negatívne komentáre, čo najrýchlejšie. Čím dlhšie ostáva komentár bez odpovede, tým viac škôd môže spôsobiť. Spoločnosti, ktoré používajú sociálne média sú stále na druhej strane len ľuďmi. Preto môžu reagovať hnevom a emóciami, ktoré ich sprevádzajú. Riešenie problému závisí od ľudskej povahy. Je potrebné sa pýtať, reagovať láskavo a adresovať odpoveď na konkrétnu osobu. Týmto spôsobom zistíme, kedy sa problém vyriešil alebo kedy je potrebné sa ospravedlniť. [9]

Ignorovať trolov, kým sa nepohnú

Trolom je niekto, kto nemá opodstatnenú sťažnosť. Je to niekto, kto chce ostatných len rozrušiť, spôsobiť chaos a negatívnu reakciu. Trolovia potrebujú pozornosť. Teóriou teda je, že ak ich budeme

ignorovať a odmietneme im venovať pozornosť jednoducho odídu a presunú sa na iný cieľ, aby získali pozornosť, po ktorej tak túžia. Na druhej strane sa môže stať aj to, že kým ich my odmietame, členovia komunity im na príspevky odpovedajú. [9]

Vytvoriť pravidlá komentovania

Každá sociálna sieť má vytvorenú politiku o spoločenských štandardoch, ale navyše si správcovia môžu vytvoriť vlastnú politiku ako prevenciu proti trolom, ktoré určujú aké správanie je, alebo nie je prijateľné. Jasné a stručné pravidlá zjednodušujú pre fanúšikov sociálneho média chápanie, ako komunita funguje a ako tam majú ako fanúšikovia fungovať. V prípade porušenia z toho môžu správcovia vyvodiť dôsledky. [7]

Počúvať

Počúvanie je dôležité pre reputáciu spoločnosti. Sociálne médiá sú v dnešnej dobe naplnené príbehmi o zlých skúsenostiach ľudí či už s leteckou spoločnosťou alebo online predajcom. Častokrát reakciou týchto spoločností je automatická odpoveď alebo dokonca žiadna odpoveď. V ešte horšom prípade spoločnosti odstraňujú takéto komentáre. Takéto skúsenosti zanechávajú v zákazníkoch veľmi zlé pocity a znižuje sa tým reputácia spoločnosti. Preto je pre spoločnosti veľmi potrebné monitorovanie reakcií zákazníkov, aby sa zistilo, čo hovoria o produktoch, či službách prostredníctvom príspevkov, priamym označovaním alebo hodnotením. [7], [10]

Opraviť chyby

V prípade, že došlo k chybe je potrebné ju opraviť a vysvetliť svojmu zákazníkovi, ktorý sťažnosť napísal, prečo vznikla. Ak je na druhej strane nespokojný zákazník, pravdepodobne ocení pripustenie chyby. Avšak ak je trolom, odpoveď neslúži pre neho ale pre komunitu. Slúži na to, aby sme sledovateľom ukázali profesionalitu. Trola môže odpoveď rozrušiť a odradiť ho od ďalšieho obťažovania. [7]

Reagovať faktami

Ak trol šíri klamlivé a falošné informácie je potrebné na to reagovať faktami. Odpovedať na takéto komentáre a opraviť dezinformáciu, ktorú zdieľa. Trola absolútne nezaujíma odpoveď, pravdepodobne veľmi dobre vie, že všetko, čo napísal je lož, skôr je táto odpoveď pre členov komunity. Ide o to, aby sa lži a klamstvá nezdieľali a nešírili ďalej. [7], [10]

Ignorovať ich

Trol chce rozrušiť ostatných a vyvolať negatívne reakcie, najjednoduchšie ako proti tomu bojovať je ignorovať ich. Trolovia majú záujem o pozornosť, teóriou teda je, že ak im tú pozornosť

odmietneme venovať, zbavíme ich sily a oni sa následne posunú ďalej, aby získali pozornosť, po ktorej túžia. Na druhej strane môže byť takýto postoj aj negatívny. Zatiaľ kým ich správcovia sociálnych médií odmietajú, členovia môžu neúmyselne naplniť ich túžbu po pozornosti tým, že budú odpovedať na ich príspevky. [7], [10]

Rozpustiť situáciu humorom

Táto stratégia je v teórii jednoduchá a efektívna ale realizácia môže byť ťažšia. Používanie humoru môže rozptýliť situáciu ale je dôležité dávať pozor na to, aby použitie humoru nebolo nerozumné. Nič nie je horšie ako vtip, ktorý vtipom vlastne nie je. [7]

Blokovať alebo zakázať

Väčšina trolov je neškodných alebo nepríjemných. Existujú však prípady, kedy trol zájde príliš ďaleko, či už vyhrážkami alebo prejavom nenávisti. V takýchto prípadoch je rozumné používateľa zablokovať alebo mu zakázať činnosť na sociálnych sieťach. [7]

Neodstraňovať príspevky

Odstránenie príspevku trola môže vyvolať ešte horšie správanie a to také, že v budúcnosti napíše ešte niečo horšie. Je potrebné si dávať pozor pri odstraňovaní príspevkov, nakoľko sa môže neúmyselne stať, že odstránime príspevok nespravodlivo používateľovi, ktorý nie je trolom. [7]

Je potrebné si pamätať, že trol je narušiteľ. Vo všeobecnosti nemá dostatočné vzdelanie na zapojenie sa do debaty. On len skrátka chce aby sme sa rozčúlili, rozhnevali, či rozrušili. Chcú prevziať diskusiu a preformulovať ju ako analýzu toho, ako zlý alebo hlúpy údajne sme. Preto je vhodné ukázať svoju pravú povahu a použitím vhodnej metódy ich vypustiť z diskusie. [7]

2.3. Falošné správy

Falošná správa je neologizmus, ktorý sa veľmi často používa na označenie vymyslenej správy. Tento typ správy pozostáva z úmyselnej dezinformácie a podvodu, ktorá sa šíri prostredníctvom tradičných tlačových a vysielaných spravodajských médií alebo online sociálnych médií. Tieto vymyslené informácie sú distribuované hlavne sociálnymi médiami ale taktiež sa pravidelne šíria aj prostredníctvom bežných médií. Falošná správa je napísaná a uverejnená s cieľom zavádzať a poškodiť dobré meno spoločnosti, subjekt alebo osoby a profitovať z toho buď finančne alebo politicky. Používajú nečestné alebo úplne vymyslené titulky v snahe zvýšiť čitateľnosť a svoj príjem zo zdieľania na internete. Záhadné zavádzajúce a klamlivé falošné správy sa líšia od paródie, ktorá má pobaviť, než zavádzať svoje publikum do omylu. [11], [12]

Význam falošných správ sa v prvom rade zvýšil v politike. V prípade médií je potrebná schopnosť prilákať divákov na ich stránky sociálnych médií, aby získali financie z online reklamy. Ak sú publikované príbehy s nepravdivým obsahom priťahuje to používateľov, čo prináša prospech danej inzercie a zlepšuje hodnotenie príbehu. Ľahko získané financie z online reklám, zvýšená politická popularizácia a popularita sociálnych médií, to všetko sa podieľa na šírení falošných správ. [11]

Falošné správy podkopávajú seriózne mediálne spoločnosti a sťažujú novinárom, aby uvádzali významné spravodajské príbehy. Analýza americkej internetovej mediálnej a spravodajskej spoločnosti BuzzFeed zistila, že top dvadsať falošných správ o prezidentských voľbách z roku 2016 v USA získalo viac angažovanosti na Facebooku ako prvých dvadsať volebných príbehov z devätnástich hlavných sociálnych médií. Taktiež známe vydavateľstvá boli anonymne napadnuté stránkami, ktoré zámerne vydávajú falošné správy, pretože je ťažké detekovať zdroje falošných správ. Počas a po prezidentských voľbách začal Donald Trump používať termín falošná správa na opisovanie negatívnej informácie o jeho predsedníctve. [11]

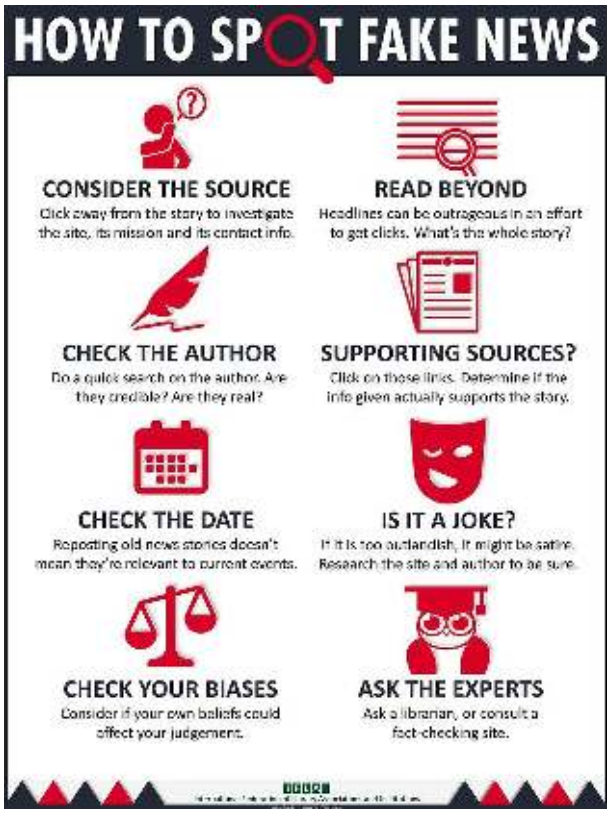
Existuje viacero typov falošných správ: [11]

1. Satira alebo paródia - žiadny úmysel pre spôsobenie škody ale je tam možné pobláznenie
2. Nepravdivá spojitosť - titulky a nadpisy nepodporujú obsah
3. Mylný obsah - zavádzajúce používanie informácií na konfrontáciu problému alebo jednotlivca
4. Falošný kontext - ak je skutočný obsah zdieľaný s falošnými kontextovými informáciami
5. Podvodný obsah - ak sú pôvodné zdroje zásobené falošnými zdrojmi
6. Manipulovaný obsah - keď sú zmanipulované originálne informácie alebo snímky určené na podvod, napríklad pri „upravenej“ fotografii
7. Vymyslený obsah - nový obsah je 100% falošný, určený na oklamanie a poškodenie

Medzinárodná federácia knižných asociácií a inštitúcií, ktorá zastupuje záujmy ľudí, ktorí sa spoliehajú na knižných a informačných pracovníkov zverejnila zhrnutie vo forme diagramu na pomoc ľuďom rozpoznať falošné správy. Obsahuje osem bodov, ktoré uvedieme nižšie a zobrazíme daný diagram: [11]

1. Zvážiť zdroj - pochopenie účelu a cieľa
2. Čítanie aj mimo nadpisu - pochopenie celého príbehu
3. Skontrolovať autorov - zistenie ich skutočnosti a dôveryhodnosti
4. Posúdenie podporných zdrojov - zabezpečenie podpory požiadavky
5. Skontrolovať dátum uverejnenia - či je obsah relevantný a aktuálny

6. Opýtať sa, či je to vtip - určenie, či to znamená satiru
7. Preskúmať svoje vlastné predsudky - zistiť, či s ovplyvňuje vlastný názor
8. Opýtať sa odborníkov - získať potvrdenie od nezávislých ľudí so znalosťami



Obr. 1 Ako spozorovať falošné správy

[11]

Snaha manipulovať mienku ľudí nie je síce rovnako stará ako ľudstvo samotné, ale delí ich len zopár generácií. Ľudstvo však muselo meniť svoju podobu podľa toho, aké spoločenské zriadenie práve bolo pri moci. V staroveku boli pri moci kmeňoví vodcovia, kniežatá, králi či faraóni. Ak niekto chcel manipulovať s mocou, stačilo ovplyvniť tých, ktorí boli práve pri moci. [12]

S príchodom mestských štátov bolo potrebné získať si širšie skupiny ľudí ako senátorov či vyslancov. Do popredia sa dostala rétorika, rečnícke umenie ako zdvorilo podať aj chúlostivé veci a získať si nadšenie publika. Na konci stredoveku postupne mizli klasické zriadenia a nahradili ich konštitučné monarchie, republiky a iné formy všeobecného ľudového spravovania. S príchodom demokratických režimov bolo potrebné o svojej pravde presvedčiť masy ľudí, k čomu prospeli

nástroje masovej komunikácie. Prejavy sa vydávali vo forme kníh, či v dennej tlači. Problémom v tom čase bolo to, že čitateľ si knihu alebo noviny s klamstvami museli kúpiť. [12]

Pre manipulátorov sa zrodila nádej pri zavedení rozhlasu a televízie. Rovnako ako tlačové média, umožňovali rozšíriť informáciu medzi masy ľudí, ale už za obsah nemuseli platiť, čo znamenalo, že mediálna manipulácia sa stala udržateľnejšou. Celonárodné rozhlasy a televízie prevádzkovali vo väčšine krajín štát, nakoľko ich prevádzka bola v tom čase nákladná. To umožňovalo štátom ušetrenie akejkoľvek konkurenčnej kritike. V krajinách východného bloku sa cenzúra rozhlasu a televízie stala nástrojom politického prežitia, ako je napríklad do dnes v Severnej Kórei. Z dnešného uhla pohľadu aj to boli falošné správy, ale zaujímavé je, že vtedy sa to za to nepovažovalo. [12]

Dôvodom, prečo bol vplyv televízie a rozhlasu tak účinný bol fakt, že neexistoval popri nich kanál, ktorý by menším skupinám ľudí, ktorí vedia, že nie je propagovaná pravda umožnil zvyšku obyvateľstva ukázať pravdu. Masová manipulácia začala pomaly opadávať s príchodom internetu. Zrazu bolo veľmi jednoduché a rýchle zistiť, čo je pravda, a čo zas nie. Televízia stratila informačný monopol a masy ľudí sa už viac nedali manipulovať. Ak nejaký vysokopostavený vplyvný človek chce manipulovať verejnou mienkou už sa nestačí postaviť do televízie a rečniť. Ktokoľvek, kto pozná problematiku môže takého človeka sfúknuť na svojom blogu alebo sociálnej sieti. Manipulácia musela tým pádom nabrať väčší rozmer. Nižšie uvedieme pár zaujímavostí od staroveku až po súčasnosť, ktoré boli spôsobené vplyvom falošných správ: [12]

Starovek

- V 13. storočí pred naším letopočtom Rameses Veľký rozšíril lži a propagandy o tom ako zvíťazili Egypťania v bitke pri Kádeš. Pritom zmluva odhalila, že bitka bola len obyčajným výmyslom.
- Počas prvého storočia pred naším letopočtom panovník Octavian uskutočnil kampaň dezinformácie o svojom súperovi Marc Antony, ktorý o ňom rozširoval klamlivé údaje a zobrazoval ho ako opilca a sukničkára. Marc Antony sa nakoniec zabil po prehratej bitke potom, ako sa dopočul, že jeho milovaná spáchala samovraždu, čo bolo len falošnou správou.
- Počas druhého a tretieho storočia nášho letopočtu sa šírili klamlivé povesti o tom, že sa kresťania podieľajú na rituálnom kanibalizme a inceste. [11]

Stredovek

- Klamlivý príbeh o kráľovi Haroldovi hovorí, že zomrel potom, ako mu šíp zasiahol oko. V skutočnosti bol v roku 1066 zastrelený.

- V roku 1475 falošný spravodajský príbeh tvrdil, že židovská komunita zavraždila kresťanské dieťa. Samotný pápež, ktorý bol v tom čase pri moci sa pokúšal vyvrátiť príbeh ale to už bolo bezúspešné, nakoľko sa táto informácia až príliš rozšírila. [11]

Skoré moderné obdobie

- Po vynájdení tlačiarne v roku 1439 sa publikácie stali rozšírenými ale neexistoval žiaden štandard novinárskej etiky, ktorý by bol jasne daný.
- V roku 1610 sa zvýšil dopyt po overených informáciách.
- Počas 18. storočia boli vydavatelia falošných správ v Holandsku pokutovaní a bola im zakázaná činnosť.
- Počas obdobia vlastnenia otrokov v Spojených štátoch podporovatelia otroctva propagovali falošné príbehy o Afroameričanoch, ktorých ľudia bielej rasy považovali za ľudí s nižším statusom. [11]

19. storočie

- V roku 1835 publikoval newyorský časopis články o veľkom mesiaci, ktoré hovorili o objavení života na Mesiaci, či civilizácie. Fiktívne články prilákali nových čitateľov. O mesiac neskôr sa priznali, že to bola séria podvodov, ktoré mali byť určené pre zábavu čitateľov a nie ich zavádzanie. [11]

20. storočie

- Počas prvej svetovej vojny sa šírili klamlivé zverstvá o tom, že padlí nemeckí vojaci boli zužitkovaní pre tuk použitý na výrobu sviečok, mazív, či nitroglycerínu. Nesprávne povesti o takejto továrni sa šírili v médiách v čase, kedy Veľká Británia presviedčala Čínu o spojenectve. Tento príbeh využili počas druhej svetovej vojny pre britskú propagandu.
- Nacisti využívali tlačovú i vysielanú žurnalistiku na propagáciu svojich programov, a to v záujem získania týchto médií alebo zvyšovania politického vplyvu. V priebehu druhej svetovej vojny boli použité falošné správy za účelom presvedčiť širokú spoločnosť doma ale aj v nepriateľských krajinách. Na druhej strane aj Briti využívali rozhlasové vysielanie a distribuované letáky pre odradenie nemeckých jednotiek. [11]

2.3.1. Falošné správy a 21. storočie

V 21. storočí sa rapídne rozšíril vplyv falošných správ spolu aj s používaním tohto výrazu. Otvorenie internetu pre ľudí v deväťdesiatych rokoch bolo obrovským posunom umožňujúci prístup k informáciám. Postupom času sa internet rozrástol do nepredstaviteľných rozmerov s množstvom informácií, ktoré sa neustále množia. Práve pre toto sa stal internet hostiteľom mnohých nežiaducich,

nepravdivých a zavádzajúcich informácií, ktoré môže zdieľať a zverejňovať ktokoľvek. Okrem odkazov na pripravené príbehy, ktoré sú zamerané na podvádzanie čitateľov, na ich klikanie, zvyšovanie návštevnosti a zisku sa termín falošné správy vzťahuje aj na satirické správy, ktorých účelom nie je zavádzať ale skôr informovať používateľov a zdieľať humorné komentáre. [11]

Hlavným cieľom sprístupnenia internetu bolo vyhľadávanie a prístup k informáciám. Keď boli na internete predstavené falošné správy, pre niektorých používateľov bolo veľmi ťažké nájsť pravdivé informácie. Vplyv falošných správ sa postupne stal svetovým fenoménom. Falošná správa sa častokrát šíri prostredníctvom falošných spravodajských webových stránok, ktoré si dôveryhodnosť získavajú tým, že sa špecializujú na vytváranie noviniek zameraných na pozornosť, ktoré sú podpísané pod známe spravodajské zdroje. Tvorca World Wide Web vyhlásil, že falošná správa je jednou z troch nových znepokojujúcich internetových trendov, ktoré musia byť vyriešené, inak internet nebude slúžiť ľudstvu. Ďalšími rušivými trendmi, ktoré ohrozujú internet sú prudké vlny používania internetu na účely dohľadu nad občanmi, ako aj na účely kybernetickej vojny. Jeden z výskumov ukázal, že falošné správy škodia sociálnym médiám a online predajniam oveľa viac ako tradičné tlačové a televízne kanály. [11]

Falošná správa 21. storočia je často zameraná na zvýšenie finančných ziskov sociálnych a spravodajských médií. V rozhovore s bývalým generálnym riaditeľom falošného mediálneho podniku vyrozprával, kto píše falošné spravodajské články, kto financuje tieto články a prečo vlastne vytvárajú a distribuujú nepravdivé informácie. Po jeho odchode uviedol, že jeho spoločnosť zamestnávala 20 až 25 spisovateľov a z reklamy získali profit od 10000 až 30000 dolárov mesačne. Do tohto biznisu vstúpil so zámerom, aby dokázal sebe aj ostatným ako rýchlo sa dokážu šíriť falošné správy. Používatelia sociálnych médií zohrávajú taktiež dôležitú úlohu pri podávaní a šírení falošných spravodajských príbehov tým, že z nich vytvoria senzačné príbehy a jednotlivci v podstate financujú tieto falošné spravodajské webové stránky a ich obsah. [11]

Mnoho fiktívnych novinových článkov na internete pochádza z mesta v Macedónsku, kde rôzne falošné spravodajské organizácie zamestnávajú stovky tínedžerov, aby, čo najrýchlejšie vytvárali a zdieľali senzačné príbehy pre rôzne americké strany a spoločnosti. [11]

Zelená spoločnosť zistila, že vzdelaní a bohatí ľudia vo veku od 40-50 rokov sú primárnymi spotrebiteľmi falošných správ a sú to ľudia, ktorí sa prikláňajú k falošným názorom. [11]

Štúdia na Oxfordskej univerzite z roku 2018 zistila, že podporovatelia Trumpa prijali najviac nevyžiadaných správ prostredníctvom sociálnych sietí Facebook a Twitter. Ďalší výskum skúmal odbyt falošných správ počas americkej prezidentskej kampane v roku 2016. Ich zistenia ukázali, že podporovatelia Trumpa a Američania, ktorí majú viac ako šesťdesiat rokov boli s určitou

pravdepodobnosťou vystavení väčšiemu počtu falošných správ ako fanúšikovia Clintonovej. Tieto nepravdivé správy, ktoré si fanúšikovia čítali boli najviac rozšírené pomocou sociálnej siete Facebook a viedli ľudí k tomu, aby navštívili klamlivé spravodajské webové stránky. [11]

Americkí výskumníci sa zaoberali rozsiahlou štúdiou, ktorá porovnávala 126 000 tweetov, ktoré viac ako tri milióny používateľov publikovalo medzi rokmi 2006 a 2017. Správy roztriedili medzi pravdivé a nepravdivé, respektíve falošné a tým vedci zistili, že výrazne rýchlejšie sa šíria práve tie falošné. Ich vyjadrením bolo, že falošnosť sa šíri oveľa rýchlejšie, hlbšie, ďalej a širšie ako pravda. Výrazný rozdiel sa zaznamenal správ týkajúcich sa politiky, než pri iných témach ako terorizmus, veda, financie, či prírodné katastrofy. Podľa danej štúdie je za rýchlosťou najmä element neznámej alebo nepreskúmanej témy, čo naznačuje, že používatelia s najväčšou pravdepodobnosťou zdieľajú neznáme informácie. Reakciou na falošné správy bol strach, znechutenie a prekvapenie na rozdiel od reakcií na pravdivé príbehy, ktoré boli radosť, dôvera, či ďalšie očakávanie. Ďalším významným zistením štúdia bol aj výsledok, kto zdieľa a šíri falošné správy. Predpokladalo sa, že za šírenie klamlivých správ stoja najmä automatizované a roboticky zdieľané počítače, ale opak bol pravdou. Najviac sú tieto správy šírené práve používateľmi. V súčasnosti je najžiadanejším spôsobom ako bojovať proti falošným správam najmä vzdelávanie a to nie len pre študentov a mladých, či na školách. Ďalej má byť podporovaná kvalitná žurnalistika. [13]

Ďalšou záležitosťou v bežných médiách je používanie filtrovaných správ, tzv. bubliny, ktorá bola vytvorená za účelom podávania konkrétnych informácií používateľovi na platformách sociálnych sietí, ktoré sa im budú páčiť. Takto sa taktiež vytvárajú falošné a zaujaté správy nakoľko sa zdieľa iba časť príbehu, ktorú si ma divák obľúbiť. Svet sa čoraz viac prispôsobuje informačným technológiám. Falošné správy sa stávajú problémom v dnešnej spoločnosti, pretože ľudia vidia len časť problémov a nie všetky problémy, čo spôsobuje, že je oveľa ťažšie tieto problémy riešiť. [11]

Overenie vierohodnosti informácií, ktoré získavame z neosobného zdroja si častokrát overujeme. Ak však správu získame od niekoho blízkeho nebudeme pátrať odkiaľ ju má. Tento efekt sa umocňuje v prípade, ak nie je informácia podaná z očí do očí, ale je zaslaná prostredníctvom sociálnych sietí a my si ju prečítame s časovým odstupom, ako nám bola doručená. Keby bola táto informácia klamlivá, tak by ju určite už niekto vyvrátil. Sociálne siete poskytujú ideálne prostredie pre nižšie uvedené predpoklady falošných správ: [12]

a) Ak sa na sociálnej sieti rozšíri nepravdivá informácia medzi najbližších priateľov je tam ešte možnosť otázok o zdroji tejto informácii. Ak sa však rozšíri už aj medzi ich priateľov buď zaslaním alebo lajkovaním niektorého z ich blízkych, nebudú kontaktovať daného spúšťača o jej pravdivosti.

b) Každý kto prepošle falošnú správu pridáva jej na sile. Ak prichádza klamlivá informácia z viacerých zdrojov v podobnom období, nie je ťažké uveriť, že je to seriózna informácia. Šírenie správ po sociálnych sieťach zapája oveľa viac používateľov práve vďaka ich kopírovaniu.
c) Väčšina používateľov sociálnych sietí im venuje len rýchly pohľad. Čas a priestor konfrontovať zdroj informácií je výrazne nízka. Častokrát len zaregistrujeme, že nejaká falošná správa prešla okolo nás.
d) Tento problém sa umocňuje tým, že veľké množstvo používateľov si neoveruje pravdivosť informácii z dôvodu, že daná informácia sa páči už tisícom používateľov. Načo by plytvali svojím časom, keď si danú informáciu overovalo už stovky ľudí pred nimi.
e) Za šírenie masového klamstva pri tlačových médiách musel platiť aj vydavateľ aj príjemca. Náklady príjemcu odpadli pri zavedení televízie, ale stále mal šíriteľ náklady na vysielanie. Náklady prijímateľa ako aj náklady vysielateľa opadli pri zavedení sociálnych sietí. Tým sa sprístupnila tvorba falošných správ a táto šanca je poskytnutá aj pomerne malým skupinám, či dokonca jednotlivcom.

Takto nadobudla éra masovej manipulácie novú podobu. Stačí mať konto na sociálnych sieťach, na ktorom sa vytvorí profesionálne vyzerajúca správa , či článok. Následne je zdieľaná medzi svojich najbližších a už postačí len čakať ako sa klamlivá správa bude šíriť rýchlo a zadarmo. [12]

2.3.2. Detekcia falošných správ na sociálnych sieťach

Detekcia falošných správ na sociálnych médiách predstavuje nové výzvy, priťahuje výskumníkov a má jedinečné vlastnosti. Falošné správy sú zámerne napísané pre zavádzanie čitateľov, aby týmto nepravdivým informáciám uverili, čo následne sťažuje ich zistenie na základe obsahu správ. Preto je potrebné zahrnúť pomocné údaje, ako sú napríklad sociálne zväzky používateľov na sociálnych médiách pre vytvorenie odlišností od skutočných správ. Využívanie týchto pomocných údajov nie je nijako triviálne a samo o sebe sociálne zväzky používateľov s falošnými správami vytvárajú dáta, ktoré sú veľké, neúplné, neštruktúrované a nápadné. Tieto stručné informácie sú založené na nedávnom prieskume, ktorý sa zaoberá problémom detekcie falošných správ na sociálnych sieťach a predstavuje najnovšie vedecké poznatky, ktoré uvedieme nižšie. [14]

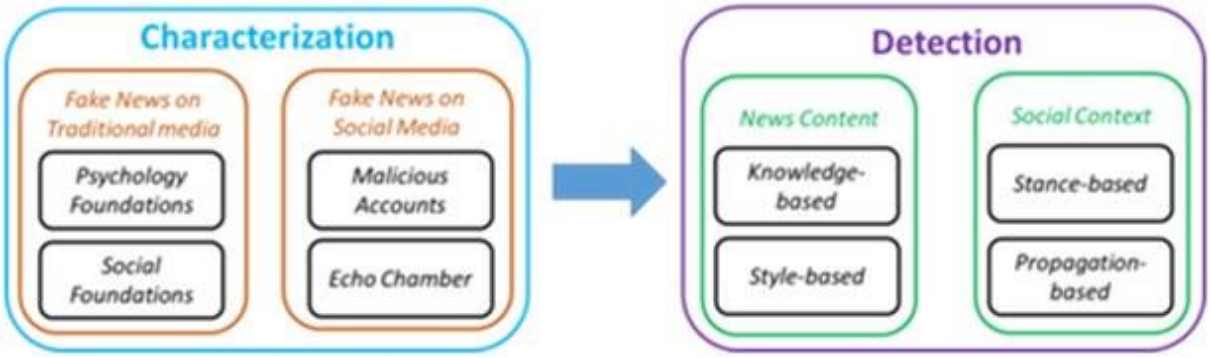
Charakterizácia a detekcia

Obrázok nižšie nám zobrazuje prehľad detekcie falošných správ na sociálnych sieťach vrátane dvoch fáz: charakterizácie a detekcie. Samotná falošná správa nie je novým problémom, média postupom času menili svoju podobu od novín cez rádio, či televíziu, až po online správy a sociálne

médiá. Vplyv klamlivých správ na tradičné médiá možno opísať z pohľadu psychológie a sociálnych teórií. Dva hlavné faktory z pohľadu psychológie robia používateľov prirodzene zraniteľných: [14]

i. Naivný realizmus - používatelia veria, že ich vnímanie reality je jediným správnym názorom.
ii. Potvrdzujúce predsudky - používatelia uprednostňujú informácie, ktoré potvrdzujú ich existujúce názory

Falošné správy na sociálnych sieťach majú svoje jedinečné vlastnosti. Napríklad škodlivé účty, ktoré sú veľmi rýchlo a ľahko vytvorené na zvýšenie šírenia falošných správ, ako sú sociálne roboty, či trolovia. Okrem toho sú používatelia vystavení určitým typom správ, kvôli spôsobu, ktorým sú zobrazené na domovskej stránke sociálnych médií. Práve z toho dôvodu majú používatelia tendenciu vytvárať skupiny, v ktorých sa združujú rovnako zmýšľajúci používatelia, kde polarizujú svoje názory. [14]



Obr. 2 Detekcia falošných správ na sociálnych sieťach od charakterizácie po detekciu [14]

Vyššie uvedené teórie sú cenným prínosom pri vedení výskumu detekcie falošných správ. Existujúce algoritmy na ich detekciu možno kategorizovať ako: [14]

i. Obsah správ - prístupy založené na obsahu správ sa zameriavajú na získanie rôznych vlastností založených na vedomostiach a štýle. Keďže klamlivé správy šíria nepravdivé tvrdenia, cieľom prístupov založených na vedomostiach je využívanie externých zdrojov pre skontrolovanie pravdivosti tvrdení novinového obsahu. Navyše, vydavatelia falošných správ majú častokrát zlomyseľné úmysly šírenia skreslených a zavádzajúcich informácií, ktoré si vyžadujú osobité štýly písania, pomocou ktorých sa odvolávajú a presviedčajú široký okruh používateľov. Tieto štýly nie sú viditeľné pri skutočných spravodajských článkov. Prístupy založené na štýle sa snažia odhaliť klamlivé správy zachytením štýlom písania.

ii. Sociálny kontext - prístupy založené na sociálnom kontexte sú zamerané na využitie používateľských sociálnych zväzkov, ktoré slúžia ako pomocné informácie pri odhaľovaní falošných správ. Prístupy založené na postoji využívajú názory používateľov z príslušného obsahu príspevkov na odvodenie pravdivosti originálnych článkov.

Online správy môžu byť zhromažďované z rôznych zdrojov, ručné určovanie pravdivosti správ je náročná úloha, ktorá vyžaduje odbornú znalosť domény, analýzu tvrdení, dodatočné dôkazy, kontext a správy z autoritatívnych zdrojov. Existujúce verejné súbory dát o falošných správach sú skôr obmedzené práve kvôli vyššie uvedeným výzvam. S cieľom uľahčiť výskum detekcie falošných správ bol vytvorený súbor dát, ktorý obsahuje novinky a sociálne kontextové vlastnosti, ktoré sú práve falošnými správami. [14]

Detekcia falošných správ na sociálnych sieťach je pomerne novo vznikajúcou oblasťou výskumu. Daný prieskum sa zaoberá súvisiacimi oblasťami výskumu, otvorenými problémami a budúcimi výskumnými smermi z perspektívy dolovania dát. Na obrázku nižšie sú znázornené výskumné smery v štyroch perspektívach. [14]

Obr. 3 Budúce smerovanie a otvorené problémy pri detekcii falošných správ na sociálnych médiách [14]

Dátové zameranie **-** zameriava sa na rozličné aspekty dát falošných správ, ako je zhromažďovanie porovnávacích údajov, psychologické overenie falošných správ a skoré odhalenie falošných správ.

Zamerané na funkcie- cieľom je preskúmať efektívne funkcie na detekciu falošných správ z viacerých zdrojov, ako je napríklad obsah správ alebo sociálny kontext.

Zamerané na modely **-** otvára dvere na vytváranie praktických a efektívnych model na detekciu falošných správ, vrátane kontrolovaných, čiastočne kontrolovaných a nekontrolovaných modelov.

Zamerané na aplikácie **-** zahŕňa výskum, ktorý presahuje detekciu falošných správ, ako falošné správy difúzie a intervencie. [14]

Uvedieme niekoľko spôsobov ako na boj s falošnými správami: [15]

1. ***Skórovanie webových stránok*** - metóda propagovaná gigantom Google. Spoločnosť berie do úvahy presnosť prezentovaných faktov, ktoré môžu vyhodnocovať webové stránky. Daná technológia nadobudla význam, nakoľko sa pokúša porozumieť kontextu stránok bez toho, aby sa spoliehali na signály od tretích strán.
2. ***Zváženie faktov*** - v boji proti klamlivým správam je nevyhnutné zvážiť fakty, ktoré sa správy snažia zdieľať. Umelá inteligencia prechádza predmet príbehu, nadpis, text, či geografické umiestnenie. Okrem iného dokáže zistiť, či iné stránky vykazujú rovnaké skutočnosti. Týmto spôsobom sa zistené fakty porovnávajú so známymi zdrojmi s využitím umelej inteligencie.
3. ***Predvídanie renomé*** **-** povesť zdroja zdieľaných správ slúži na to, aby sa zablokovali novinky falošných správ. Je potrebné nevzbudiť pochybnosť o povesti zdroja informácií. Momentálne je možné určiť pravosť webových stránok pomocou modelu strojového učenia, ktorý dokáže predpovedať reputáciu týchto stránok.
4. ***Objavenie senzačných slov*** **-** pokiaľ ide o spravodajské články, nadpis slúži pre zachytenie pozornosti publika. Práve preto sa senzačné titulky stávajú užitočným nástrojom pre vytvorenie záujmu čitateľa. Keď sa používajú senzačné slová na šírenie falošných správ, láka to viac očí a tým sa správy šíria širšie a rýchlejšie. Momentálne sa bojuje proti tomuto pomocou nástrojov umelej inteligencie, ktoré objavujú a označujú falošné spravodajské tituly pomocou analýzy kľúčových slov.

Spoločnosť Facebook je jedným z najväčších obetných baránkov šírenia falošných správ, ktoré sa rapídne zrýchlilo po amerických voľbách. Stránky sociálnych médií, ako sú Facebook a Twitter, sú nepochybne hlavnými kanálmi na šírenie klamlivých informácií. Reklamná platforma Facebooku uľahčuje tvorcom falošných správ, aby svoje dezinformácie rozširovali s cieľom zvýšenia zisku práve vďaka zobrazovaniu reklám za pomoci týchto falošných článkov. V dôsledku toho sa spoločnosť

Facebook pokúsila implementovať nástroje na detekciu a boj proti falošným správam. Prvým riešením je, že používatelia majú možnosť označiť položku, ktorú považujú za falošnú. Na identifikovanie zdroja sú vytvorené odznaky, ktoré označujú klamstvo a umožňuje to používateľom získať viac informácií o príbehu. Keď dostatočné množstvo používateľov označia príbeh za falošný, znižuje sa frekvencia zdieľaného článku. Počas testovania tohto nástroja sa zistilo, že odznaky neúmyselne ukryli informácie, ktoré vysvetľovali nejasnosti a nepresnosti danej informácie. Facebook považuje falošnú správu za finančnú motiváciu, ktorá sa zameriava na kliknutia, ktoré používateľov presmerujú na webové stránky obsahujúce prevažne reklamy. Za prevenciu šírenia falošných správ spoločnosť znižuje počet označených príspevkov a tým znižuje ich návštevnosť. Opakovaní páchatelia, ktorí šíria klamlivé správy majú reklamné práva odstránené, tým sa znižuje ich distribúcia a zároveň aj možnosť zárobku. Ďalším nástrojom je implementovanie umelej inteligencie na odhalenie falošných správ. Umelá inteligencia sa dokáže rýchlo a efektívne učiť. Používa metódu pokus - omyl, pri ktorej sa zisťuje, ktoré slová, slovné spojenia a vety majú najväčší ohlas. A tie následne kombinuje tak, aby sa dosiahla maximálna emocionálna reakcia. Nejde jej o pravdu či posolstvo používateľa. Ide jej o počet reakcií, ktorá pritiahne najväčšiu pozornosť a následne ju bude šíriť ďalej. Cieľom je šokovať, preháňať, urážať, robiť všetko preto aby to v používateľoch vyvolalo potrebu reagovať. Pritom ani používatelia netušia, že majú dočinenia so strojom. Zakladateľ spoločnosti Facebooku si zobral inšpiráciu od vyhľadávača Google. Práve on je najpopulárnejším vyhľadávačom na internete. Ďalej je najpresnejší, na diskových poliach je uložených najväčší počet webových stránok, dokáže pochopiť zámer otázky používateľa a jeho očakávanie. Prečo je tak obľúbený je fakt, že je najdôveryhodnejší. Manipulácia s poradím vyhľadávania je vylúčená a jeho algoritmus funguje bez zásahu človeka. Práve z týchto a ďalších dôvodov začala spoločnosť Facebook využívať umelú inteligenciu na identifikáciu nevhodných príspevkov. Tá rozpoznáva extrémizmus, násilie, nenávisť, vyhrážky a iné formy nevhodných príspevkov. Pri automatickej detekcii falošných správ sa inšpirovali už existujúcou prácou, v ktorej umelá inteligencia dokáže predpovedať koľkokrát bude nejaký príspevok zdieľaný, koľko bude mať lajkov a komentárov. Tým istým spôsobom sa dokáže predpovedať, či príspevok bude nahlásený alebo nie. Zakladateľ spoločnosti predpovedá, že podobný algoritmus implementovaný na detekciu falošných správ bude funkčný za pár rokov. Dôvodom je fakt, že je potrebné nazbierať obrovské množstvo správ označených používateľmi alebo expertmi, aby algoritmus fungoval s nejakou štatistickou chybou a to bude niekoľko rokov trvať. Momentálne majú novinári, aktivisti a organizácie na starosti označovanie falošných správ. [16], [17], [18], [19], [22]

Za šírením falošných správ je z časti aj psychológia. Ešte z čias Tretej ríše bol povedaný výrok, že stokrát opakovaná lož sa stane pravdou. Vedci v súčasnosti zistili, že skutočne sa často opakované fakty ľahšie pamätajú. A to, čo si pamätáme, máme tendenciu označiť častokrát za pravdu. A to

dokonca aj vtedy, keď už na začiatku vieme, že je to klamstvo. Ďalším psychologickým efektom je kognitívna disonancia, ktorá taktiež uľahčuje šírenie klamlivých správ. Používateľ zavrhne akýkoľvek fakt, s ktorým nie je presvedčení alebo mu neverí. Je to pre človeka príliš nepohodlné, keď v sebe drží protichodné fakty. Americký historik vedy radí: [19]

- nezapájať do diskusie emócie,
- je potrebné diskutovať, a nie útočiť,
- pozorne počúvať a pochopiť argumenty druhej strany,
- rešpektovať svojho oponenta,
- porozumieť, prečo má oponent daný názor,
- prijatie iného, nového názoru neznamená zmenu svojho presvedčenia.

Existuje ešte robustné množstvo nástrojov, ktoré sa ukázali ako užitočné pri odstraňovaní a detekcii falošných správ. Forbes.com uviedol súhrn niektorých najpoužívanejších nástrojov na boj proti falošným správam: [15], [20]

Spike - umožňuje identifikovať a predpovedať rozprávkové príbehy, ako aj vírusové príbehy. Analyzuje hory údajov zo sveta správ.

Hoaxy - pomáha užívateľom identifikovať falošné spravodajské stránky.

Snopes - webová stránka pomáhajúca rozpoznať falošné príbehy.

CrowdTangle - nástroj umožňujúci sledovanie obsahu a včas odhaliť sociálny obsah.

Check - pomáha overiť online porušenie správ.

Google Trends - hodnotu dokazuje strážením vyhľadávania.

Le Decodex - databáza obsahujúca webové stránky označené fake alebo real, podľa pravdivosti daných stránok.

Pheme - spoločnosť, ktorá urobila technologický skok. Dokáže prečítať pravdivosť obsahu vytvoreného používateľom a online obsahu.

Ďalším nástrojom je webová stránka areyoufakenews.com, ktorú vytvoril Estela nakoľko sa už nedokázal pozerať na falošné správy, ktoré zdieľali jeho priatelia a rodina a zahlcujú ho na sociálnej sieti Facebook. Cieľom tohto nástroja je klasifikovať zaujatosť v médiách v reálnom čase. Pomocou označených údajov vytvorených prostredníctvom open source projektov, ktoré obsahujú tisícky spravodajských zdrojov, skenuje a analyzuje ich, aby zistil, či je zadaná webová stránka zaujatá, obsahujúca falošné správy alebo je spravodlivým spravodajským zdrojom. Tento nástroj zhromažďuje

desiatky tisíc overených článkov zo spravodajských zdrojov a učí vlastnú neurónovú sieť s cieľom modelovať a charakterizovať dané články. [21]

Najideálnejším spôsobom, ako bojovať proti šíreniu falošných správ môže závisieť od používateľov, čiže od ľudí samotných. Spoločenské dôsledky falošných správ, politická polarizácia, narušenie dôvery bežných médií, či vlády sú významné. Keby väčšina používateľov vedelo, ako vysoké sú podiely viny, boli by oveľa viac opatrnejší. Keď sa objaví nový príspevok, ktorý šíri hnev, je potrebné najprv preskúmať dôveryhodnosť informácií a až tak ho šíriť ďalej. To nám dáva najavo, že nie len sociálne médiá sú výhradne zodpovedné za prevenciu, detekciu a šírenie falošných správ. Veľkú zodpovednosť zohrávajú skutočne aj používatelia. Existuje už množstvo nástrojov, ktoré sme opísali na detekciu falošných správ a stránok. Problém je však v tom, že používatelia nevykonávajú kontroly. Preto kým sa umelá inteligencia a nástroje zameriavajú na detekciu klamlivých informácií, my sa musíme sústrediť na vzdelávanie ľudí, už od detstva, aby boli schopní byť kritickými mysliteľmi a nebrať každý príbeh vážne. [22]

3. Analýza súčasného stavu

Falošné správy majú v posledných rokoch výrazný vplyv na spoločenský život a to najmä v politickom svete. Sú vážnym problémom, ktorý zhoršuje fakt rýchleho napredovania počítačových technológií, ktoré zjednodušujú proces ich vytvárania a šírenia. Detekcia falošných správ sa stáva jednou z najvýznamnejších oblastí výskumu, aj keď je jednoduchšie sledovať vplyv falošných správ ako ich detegovať. [24]

Nasleduje opis niekoľkých prípadových štúdií, ktoré sa zaoberali detekciou falošných správ z rôznych uhlov pohľadu.

3.1. Prípadová štúdia 1

Prvou prípadovou štúdiou je článok *Fake Review Detection via Exploitation of Spam Indicators and Reviewer Behavior Characteristics*, ktorý zavádza niekoľko prístupov na odhalenie falošných recenzií a aktivitu ich recenzentov. Navrhovaný model využíva prehľad produktov, ako aj charakteristické správanie autorov. Skúma recenzie vytvorené počas „podozrivých“ časových intervalov. Navrhovaný prístup je potvrdený skutočným súborom recenzií spoločnosti Amazon. Výsledky experimentu dokazujú, že táto metóda úspešne identifikuje spamové recenzie vďaka použitým technikám a indikátorom. Elektronický obchod sa ovplyvnil rýchlym šírením webových a internetových technológií, ktoré umožnili zdieľanie obrovského množstva informácií vytvorených používateľmi. Spotrebitelia verejne a nepretržite zdieľajú svoje názory na zakúpené výrobky alebo ponúkané služby a posudzujú ich kvalitu. Štúdia dokázala, že online recenzie sú pre potencionálnych zákazníkov veľmi dôležité, nakoľko začleňujú online recenzie do svojho rozhodnutia. Okrem toho, spotrebitelia momentálne dôverujú online recenziám rovnako ako osobným odporúčaniam. Tento online vplyv otvoril možnosti aktivitám, ktorých cieľom je buď pravdivo zhodnotiť alebo zmanipulovať používateľské recenzie pre konkrétne služby a produkty. Veľký rozsah klamlivých recenzií sa ukázal ako veľký problém, ktorý v dnešnej dobe priťahuje záujem vedcov. Väčšinou sa zameriavajú na zlepšenie detekcie falošných správ. Detekcia falošných recenzií sa spočiatku zamerala na zdvojený obsah recenzie a súvislosť recenzie. Takáto analýza textu sa väčšinou zakladala na klasifikátoroch strojového učenia. V tejto štúdií sú skúmané najdôležitejšie ukazovatele spamu týkajúce sa kontroly a využívania charakteristík správania recenzenta, ktoré sa využívajú na klasifikáciu recenzií do dvoch tried a to buď klamlivej alebo pravdivej. Cieľom je udržať podstatnú mieru informácií, ktoré sa využívajú na identifikáciu prichádzajúcich recenzií a na sledovanie aktivity recenzentov. Navyše sa aj história skúmania podieľa na získaní ďalších údajov o všeobecnej reputácií recenzenta, ktorá pomáha určiť jeho skutočnú povahu. [23]

V priebehu posledného desaťročia sa uskutočnil podstatný výskum v oblasti spamového rozpoznávania online recenzií: [23]

Textová analýza - identifikácia falošných recenzií bola skúmaná ako úloha zistenia duplicitného textu recenzie. Duplicita obsahu bola bežne spätá s bežnou praxou spamerov, v ktorej sa rovnaká recenzia opakovane opakuje. Podobnosť medzi obsahmi recenzie je často navrhovaná ako účinná detekčná funkcia. Tieto duplicitné recenzie slúžia na klasifikáciu obsahu.

Grafické prístupy - niektoré štúdie navrhli heterogénne grafické zobrazenie medzi recenzentmi a online predajcami s cieľom zistenia nezrovnalostí. Pomocou týchto prepojení je možné určiť dôveryhodnosť recenzentov, čestnosť recenzií a spoľahlivosť predajcov.

Zisťovanie vzoru zhluku - pokiaľ sa jedná o štúdiu spamu, štúdie sa zameriavajú na aspekt času. Vzhľadom na to, že väčšina recenzentov vytvorí len jednu recenziu na daný produkt. Jedna zo štúdií vytvorila multidimenzionálny časový rad pre každý produkt na základe priemerného hodnotenia, celkový počet recenzií ku pomeru jednotlivých recenzií. Ďalšia štúdia potvrdila, že recenzie a recenzenti, ktoré sa objavujú v rovnakom zhluku posudzovanej aktivity produktu spolu často súvisia a preto pomocou grafického znázornenia modelových prepojení autorov úspešne identifikovali spamové správy. Analyzovali a posudzovali len tie recenzie, ktoré padli do zhluku časového intervalu, nakoľko práve tie obsahovali s najväčšou pravdepodobnosťou podozrivé aktivity.

Hodnotenie manipulačnej analýzy - spameri sa pokúšajú propagovať alebo redukovať produkt tým, že manipulujú s jeho celkovým poradím. Značný počet skorých hodnotení ako aj extrémnych hodnotení súvisia s podozrivým správaním. Spameri narúšajú distribúciu recenzií, čo zanecháva stopy, ktoré môžu byť použité na pomoc pri objavovaní spamových recenzentov.

Zisťovanie skupinových spamerov - klamlivý recenzenti častokrát spolupracujú navzájom, aby podporovali alebo znižovali určitý produkt alebo službu. Autori aplikovali časté metódy dolovania recenzií spoločnosti Amazon, aby rozdelili a zaradili skupiny podľa pravdepodobnosti spamovania. Novší prístup použil vzájomné zhluky spamových vzťahov na modelovanie siete, ktorá úspešne zistila skupiny spamerov.

Navrhovaná metóda preklenuje existujúce medzery zavedením efektívneho detekčného modelu falošných recenzií, ktorý pracuje na produktovej úrovni, využíva zisťovanie vzoru zhluku, detekciu podozrivých časových intervalov, ako ďalšiu analýzu a integruje minulú i súčasnú činnosť recenzenta. [23]

Tento projekt sa pokúša vytvoriť robustný systém detekcie falošných recenzií, a to práve zvážením rôznych osvedčených a akceptovaných funkcií spamu. Daný model prijíma na vstupe

množinu recenzentov spojených s výrobkom. Pre každú kontrolu sa extrahujú potrebné informácie a metadáta vrátane textu recenzie, hodnotenie recenzie, časovú značku a ID recenzenta. Používa sa aj vyhľadávanie vzoru zhluku ako doplnkový analytický nástroj na identifikáciu časových intervalov a určenie recenzií, ktoré sú podozrivé. Táto metóda zvažuje všetky recenzie produktu, bez straty informácií, zohľadňuje históriu činnosti recenzenta, nakoľko môže ovplyvniť celkovú reputáciu pri následnom určovaní recenzenta, či je klamlivý alebo čestný. Pri analýze recenzie sa skúmajú jej viaceré funkcie a charakteristiky správania ako dodatočné opatrenie dôveryhodnosti recenzenta. Hodnota spamovej kontroly sa určuje použitím funkcie lineárneho váženia hodnoty a definície prahu skóre spamu, na základe ktorého sa porovnáva skóre každej recenzie. Falošná recenzia je tá, ktorá presahuje skóre prahu a čestné recenzie sú tie, ktorých skóre tento prah spamu neprekračuje. [23]

Základné indikátory spamu

V tomto modeli sú použité tri základné indikátory spamu: [23]

Hodnotiaca odchýlka - kontrola spamu sa zvyčajne zameriava na zvýšenie alebo zníženie celkovej pozície výrobku tým, že manipuluje s jeho priemerným skóre tak, že sa jeho skóre posunie smerom k určitému smeru a následne sa odchýli od priemeru.

Počet recenzií - spameri bežne vytvárajú viacero recenzií na ten istý produkt, aby mal väčší vplyv na verejnú mienku a manipulovať s priemerným hodnotením.

Podobnosť obsahu - spameri častokrát opakujú rovnaký text recenzie, nakoľko nový autorský obsah by bo časovo náročný. Z tohto dôvodu vieme odhaliť spamerov tým, že berieme do úvahy celkovú podobnosť ich recenzií. Na tento účel sa využíva kosínusová podobnosť.

Zisťovanie vzoriek zhluku

Spameri zvyčajne vytvárajú veľké množstvo recenzií v primerane krátkom časovom období, časových intervaloch, aby mohli rýchlo negovať čestné názory a ovládnuť ich. Takéto nadmerné pridávanie má za následok náhly nárast aktivity pri skúmanom produkte, vytváranie zhlukov alebo vrchol v určitých časových intervaloch. Táto metóda skúma všetky recenzie produktu pri existencii spamu a podrobnejšie analyzuje tie, ktoré boli vytvorené v časových intervaloch. Recenzie, ktoré patria do intervalu zhluku sú extrahované a následne na nich použité dva nasledujúce indikátory spamu: [23]

Podobnosť obsahu zhluku - vysoké skóre podobnosti medzi recenziami a inými recenziami toho istého zhluku by mohlo naznačovať, že recenzia podozrivo pripomína iné recenzie. Za predpokladu, že skóre podobnosti považujeme za normálne, je možné upraviť metriku, aby boli ovplyvnené len tie recenzie,

ktoré vykazujú vyššiu podobnosť ako tie normálne, aby nedošlo k znevýhodneniu recenzií, ktoré boli pridané v zhluku časového intervalu.

Aktivita zhluku - spamer vytvorí veľké množstvo recenzií v malých zhlukoch svojej činnosti, aby rýchlo zmanipuloval všeobecný názor. Predpokladom je, že čestný recenzent vytvorí maximálne dve recenzie.

Povesť recenzie

Existujú dostatočne dostupné informácie o predchádzajúcej činnosti recenzenta, ktorá by tomuto modelu umožnila lepšie zhodnotiť celkovú reputáciu recenzenta a dôveryhodnosť skúmania recenzií. Daný model využíva tri indikátory spamu založených na histórií recenzentov: [23]

Extrémne hodnotenie - väčšina spamerov využíva extrémne hodnotenie, buď 1 alebo 5 na 5-stupňovej hodnotiacej tabuľky, s cieľom rýchlo zvýšiť alebo znížiť priemerné skóre produktu.

Počet recenzií produktu - v dôsledku nadmerného skúmania sa zvažuje relevantné správanie recenzenta na minulé recenzované produkty. Pre tento účel sa meria priemerný počet recenzií, ktoré recenzent píše na produkt tým, že rozdeľuje veľkosť historickej recenzie na počet posudzovaných produktov.

Recenzia zhlukovania - spameri majú tendenciu vytvárať všetky svoje recenzie vo veľkom množstve a v krátkom čase, zhluku, aby mohli rýchlo dominovať čestným recenziám.

Ak sa vezmú do úvahy vyššie uvedené indikátory spamu založených na histórií, vytvoríme kombinovanú metódu, ktorá modeluje dôveryhodnosť alebo povesť recenzenta. Nízke skóre naznačuje dobrú povesť, zatiaľ čo vysoké skóre znamená podozrivé správanie. [23]

Skórovacia funkcia spamu

Funkcia lineárneho váženia spája individuálne skóre generované každým indikátorom a následne je vygenerované celkové skóre spamu pre každú recenziu. Extrémne hodnotenie sa považuje za najslabší indikátor, nakoľko aj čestný recenzent môže využiť extrémne hodnotenie. Dva indikátory spamu, počet recenzií a počet recenzií produktu, ktoré súvisia s nadmerným preskúmaním, dosahujú relatívne nízku hodnotu. Podobnosť obsahu je indikátorom, ktorý najviac ovplyvňuje váhu hodnotenia. Definovaný prah oddeľuje falošné recenzie od tých skutočných. Po preskúmaní hodnôt sa prahová hodnota nastavila na hodnotu tri. Preto sú recenzie so skóre spamu prekračujúcimi prah označené ako falošné, zatiaľ čo recenzie so skóre spamu nižším ako je prah sú považované za čestné. [23]

Záverom daného projektu je navrhnutie nového prístupu na zistenie falošných recenzií. Využívajú veľké množstvo rôznych indikátorov nevyžiadanej pošty na úrovni produktu vzhľadom na recenziu a správanie recenzenta, aby sa zhromaždili a využili všetky dostupné informácie. Navyše daný model ponúka analytické funkcie založené na rozpoznávanie zhluku, ktoré umožňujú identifikáciu podozrivých časových intervalov a recenzií. Nakoniec je meraná reputácia recenzenta tým, že je skúmaná jeho história predošlých recenzií a aktivít pre lepšie určenie pravosti ich najnovších recenzií. Hodnotenie navrhovanej metódy bolo vykonané na súbore recenzií produktov spoločnosti Amazon a výsledky určili, že kombinovaná metóda je účinná pri odhaľovaní falošných recenzií. [23]

3.2. Prípadová štúdia 2

Druhou prípadovou štúdiou je projekt dátového vedca *Yunus Genes*, ktorý vytvoril webovú stránku pre odhalenie falošných spravodajských článkov. Zhromaždil viac ako 200 000 článkov, ktoré filtroval podľa témy a dátumu. Nakoniec vyselektoval 52 000 článkov z oblasti obchodu, politiky, amerických správ a sveta, ktoré boli publikované v rokoch 2016 a 2017. Z nich bolo 12 000 označených ako falošné a 40 000 bolo pravdivých. [25]

Na detekciu využil algoritmy strojového učenia, konkrétne logistickú regresiu, Random forest a XGBoost, pri ktorých sledoval presnosť daných modelov. Po vektorizácii dokumentu pomocou TF-IDF získal nasledujúce výsledky: [25]

- Random forest 82%,
- XGBoost 65%,
- logistická regresia 95%.

V závere vytvoril webovú stránku, ktorá za pomoci daných výsledkov vie vyhodnotiť vložený článok. Výsledkom je percentuálny pomer, na koľko percent je článok falošný a na koľko je pravdivý. [25]

3.3. Prípadová štúdia 3

Tretiu prípadovú štúdiu sme čerpali z článku *Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques*, ktorý uvádza prístup analýzy textu založeného na N-gramovom modely a technikách strojového učenia. Študovali a porovnali šesť rôznych klasifikačných techník a to, K-najbližších susedov (KNN), Suport Vector Machine (SVM), logistickú regresiu (LR), Linear Suport Vector Machine (LSVM), rozhodovacie stromy (DT) a Stochastic Gradient Descent (SGD), pri ktorých sa sledovala ich presnosť. [24]

V danom projekte použili niekoľko základných N-gramových znakov na základe slov a skúmali vplyv N-gramovej dĺžky na presnosť študovaných klasifikačných algoritmov. Pre zmenšenie veľkosti textovej funkcie si vybrali dve metódy, TF a TF-IDF. Dáta, s ktorými pracovali, si zhromaždili samostatne z voľne dostupných článkov. Zozbierali 12 600 falošných a 12 600 pravdivých článkov zameraných na politickú situáciu z roku 2016. [24]

Z experimentov bolo zistené, že lineárne modely získali lepšie výsledky ako tie nelineárne. Ďalej s nárastom N-gramu sa presnosť algoritmov znižuje. Najvyššia presnosť bola dosiahnutá použitím algoritmu SVM a to 92%, najnižšia presnosť bola 47,2% pomocou algoritmu KNN. [24]

3.4. Prípadová štúdia 4

Nasledujúca prípadová štúdia sa venovala odhaleniu falošných recenzií tým, že simulovali spam recenzie, vybudovali trénovaciu množinu a následne ju klasifikovali pomocou Naive Bayes klasifikátora a náhodných lesov pre sledovanie presnosti daných modelov. [26]

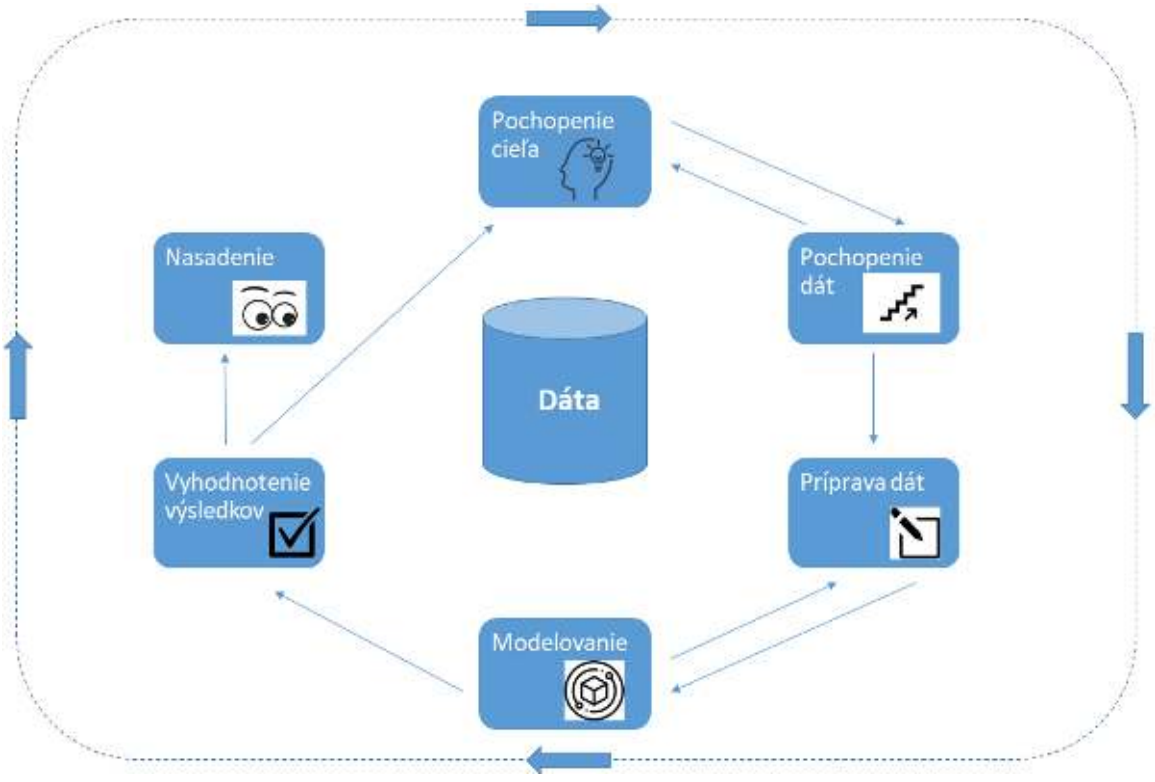
Recenzie boli získané zo spoločnosti Amazon a obsahovali webovú stránku predávajúceho, názov produktu, hodnotenie, identifikačné číslo recenzenta, tému recenzie, obsah recenzie, dátum pridania recenzie, vplyv (koľko ľudí ju považuje za užitočnú) a informáciu, či je nákup overený alebo nie. Frekvencia termínov a frekvencia recenzie používateľa mali vplyv na klasifikačné modely. Experimenty ukázali, že model náhodného stromu získal lepšie výsledky ako Naive Bayes klasifikátor. [26]

4. Praktická časť

V praktickej časti diplomovej práce sa venujeme detekcii falošných správ pomocou algoritmov strojového učenia. K dispozícii máme vzorky dát, ktoré obsahujú falošné a reálne správy. Pre riešenie daného problému je zvolená metodológia CRISP-DM, pomocou ktorej postupujeme v rámci celej praktickej časti diplomovej práce. Nástrojom, v ktorom pracujeme je programovací jazyk R, kde sú dáta pripravené a následne vytvorené modely, ktoré sú v závere vyhodnotené pomocou viacerých sledovaných kritérií.

4.1. Metodológia CRISP-DM

Metodológia CRISP-DM je štandardný procesný model, ktorý opisuje bežné prístupy používané na dolovanie dát. Daný model predstavuje sled udalostí, medzi ktorými existujú vzťahy. V praxi sa mnohé úlohy vykonávajú v inom poradí a častokrát je potrebné vrátiť sa k predošlým úlohám a zopakovať isté činnosti. Cyklus pozostáva zo 6 fáz, kde výsledok dosiahnutý v jednej fáze ovplyvňuje nasledujúce kroky. [27] Obrázok nižšie nám zobrazuje proces danej metodológie:



Obr. 4 Metodológia CRISP-DM

Popis jednotlivých fáz: [27]

POCHOPENIE CIEĽA

Prvotný krok sa zaoberá pochopením cieľov a požiadaviek z podnikateľskej perspektívy. Cieľom je odhaliť dôležité faktory, ktoré môžu projekt ovplyvniť, ako rôzne riziká alebo obmedzenia. Zanedbanie tohto kroku môže znamenať vytvorenie správnych odpovedí na nesprávne otázky. V danej fáze je potrebné:

- *stanovenie cieľov a kritérií úspešnosti*, ktoré budú sledované pri určení úspešnosti projektu z obchodného hľadiska,
- *zhodnotenie súčasnej situácie*, pomocou ktorej je možné odhaliť faktory, prostriedky alebo obmedzenia, ktoré by mohli ovplyvniť celkový výsledok projektu,
- *určenie cieľov a kritérií dolovania dát*, ktoré definujú úspešnosť projektu z technického hľadiska,
- *vytvorenie plánu projektu*, ktorý slúži na splnenie cieľa dolovania dát, čo vedie k splneniu vopred stanovených podnikateľských cieľov.

POCHOPENIE DÁT

Druhá fáza opisuje proces získania potrebných dát a taktiež o nich poskytuje základné informácie. V danej fáze je potrebné:

- *zozbieranie dát*, obsahuje popis dát a metódy, ktoré boli použité pri ich získavaní,
- *popísanie dát*, slúži pre vizualizáciu dát a opis ich vlastností,
- *preskúmanie dát*, zisťuje vzťahy medzi atribútmi,
- *zistenie kvality dát*, slúži pre zistenie chýbajúcich hodnôt, úplnosti dát a možného výskytu chýb.

PRÍPRAVA DÁT

Fáza, ktorá je potrebná pre vytvorenie správnej množiny dát vhodnej pre modelovanie. Dáta musia byť v správnom formáte. V tejto fáze sú potrebné nasledujúce kroky:

- *výber dát*, s ktorými sa bude následne vykonávať analýza a modelovanie,
- *čistenie dát*, ktoré slúži pre ich skvalitnenie,
- *konštrukcia dát*, pri ktorej napríklad dochádza ku transformovaniu hodnôt a generovaniu nových záznamov,
- *integrácia dát*, ktorá slúži pre zlučovanie dát z viacerých zdrojov.

MODELOVANIE

V tejto fáze dochádza k výberu najvhodnejších a najlepších techník modelovania. Je veľmi potrebná interakcia s fázou prípravy dát. Táto fáza zahŕňa nasledujúce kroky:

- *výber modelovacích techník*, kde je dôležité vykonávať techniky oddelene,
- *generovanie návrhu testovania*, ktorý slúži pre testovanie kvality a presnosti modelu,
- *vytvorenie modelu*, ktoré je založené na aplikácií techník na pripravenú množinu dát,
- *hodnotenie modelu*, ktoré sa vykonáva pomocou vopred určených hodnotiacich kritérií.

VYHODNOTENIE VÝSLEDKOV

V rámci danej fázy sa hodnotí miera splnenia stanovených obchodných cieľov, kde je potrebné taktiež zistiť existenciu nedostatočného úspechu modelu. Proces sa hodnotí z hľadiska kvality a splnenia pôvodných cieľov. Určujú sa taktiež neúspešné úlohy, ktoré neboli počas procesu vykonané. Na záver sa rozhodne buď o ukončení projektu alebo sa prechádza do poslednej fázy nasadenia.

NASADENIE

Poslednou fázou je nasadenie, ktoré sa vykonáva po rozhodnutí v predchádzajúcej fáze. V tejto časti je nevyhnutné implementovať výsledky do podoby použiteľnej pre zákazníka a vytvoriť plán nasadenia. Výsledkom danej fázy je vytvorenie záverečnej práce, ktorá obsahuje zhrnutie celého projektu a všetkých dosiahnutých výsledkov. Obsahuje taktiež odchýlky a časti, ktoré by bolo vhodné zlepšiť.

4.2. Metódy strojového učenia

Strojové učenie, anglicky *Machine Learning*, je podoblasť umelej inteligencie, ktorá sa zaoberá metódami a algoritmami, ktoré umožňujú programu učiť sa a následne dokázať rozpoznať zložité vzory a robiť inteligentné rozhodnutia založené na základe informácií, ktoré sa naučili. [31]

Algoritmy strojového učenia sa delia do troch kategórií podľa spôsobu učenia a podľa riešeného problému [31]:

- Učenie s učiteľom - klasifikácia a regresia
- Učenie bez učiteľa - zhlukovanie
- Učenie s posilňovaním

Pre riešenie problému danej diplomovej práce boli vybrané algoritmy strojového učenia s učiteľom, konkrétne *Naive Bayes, Rozhodovací strom, Náhodný les a Podporné vektory*. Práve tieto metódy sú najvhodnejšie pri riešení problému klasifikácie. Vybrali sme ich z viacerých dôvodov: sú najspoľahlivejšie, najzrozumiteľnejšie, najviac používané a jednoducho sa implementujú. V programovacom jazyku R, v ktorom pracujeme, sú vytvorené balíčky, ktoré jednoducho aplikujú dané algoritmy. V podkapitolách nižšie sú teoreticky opísané a taktiež sú tam uvedené jednotlivé balíčky pre vytvorenie modelov.

4.2.1. Naive Bayes

Naive Bayes je založený na štatistickom a pravdepodobnostnom prístupe. Zaoberá sa metódou klasifikácie a svojimi princípmi sa radí medzi najjednoduchšie a najzrozumiteľnejšie klasifikátory nakoľko nepracuje so zložitými schémami a výpočtami. Je založený na tzv. Bayesovej vete, ktorú, ako aj samotný algoritmus, sformuloval anglický štatistik, duchovný a filozof Thomas Bayes, po ktorom sú aj pomenované. Je pomenovaný „naive“, čiže naivný, z dôvodu, že predpokladá nezávislosť atribútov, v našom prípade slov, čo v reálnych podmienkach splnené nie je. [31]

Algoritmus Naive Bayes sa teda počíta pomocou Naivného Bayesovského klasifikátora, pravdepodobnostného klasifikátora, ktorý dokáže klasifikovať dáta pomocou výpočtov pravdepodobnosti. [31] Matematicky je Naivný Bayesov klasifikátor vyjadrený rovnicou:

$$\boldsymbol{P}(\boldsymbol{H}|\boldsymbol{X}) = \frac{\boldsymbol{P}(\boldsymbol{X}|\boldsymbol{H})\boldsymbol{P}(\boldsymbol{H})}{\boldsymbol{P}(\boldsymbol{X})}$$

- H - hypotéza,
- X - dôkaz,
- P(H|X) - výsledná pravdepodobnosť, *H* je hypotéza a *X* slúži ako dôkaz, že hypotéza *H* platí,
- P(X|H) - pravdepodobnosť dôkazu *X*, keď *H* je pravdivá hypotéza
- P(H) - pravdepodobnosť hypotézy *H*, bez ohľadu na dôkaz *X*,
- P(X) - pravdepodobnosť dôkazu *X*, bez ohľadu na hypotézu *H*.

Pri vytváraní daného algoritmu v prostredí R nám pomohol balíček *naivebayes*, ktorý slúži na jeho výpočet.

4.2.2. Rozhodovací strom

Rozhodovacie stromy, anglicky *Decision trees*, sú jedným z najstarších a najpoužívanejších algoritmov, ktoré sa využívajú na klasifikáciu a predikciu. Sú priehľadné a jednoducho interpretované, čiže užívateľom umožňuje rýchle a ľahké vyhodnotenie dosiahnutých výsledkov. Cieľom rozhodovacích stromov je na základe prvkov z trénovacej množiny, ktoré sú vstupnými údajmi, predpovedať hodnotu cieľovej premennej. Jeho štruktúra je grafická vo forme stromu, konštruovaná spôsobom zhora-nadol, ktorý sa skladá z hrán a uzlov. Uzly sú triedami alebo testovacím atribútom a hrany sú hodnotami testovacieho atribútu. Pre výber testovacieho atribútu sa využíva informačná entropia, tzv. informačný zisk, ktorý je vyjadrený nasledujúcim vzorcom [31]:

$$\mathbf{Info(D)} = -\sum_{\mathbf{i=1}}^{\mathbf{m}} \mathbf{p_i} \log_{\mathbf{2}}(\mathbf{p_i})$$

- p_i - nenulová pravdepodobnosť triedy *i* trénovacej množiny

Počas neskorých sedemdesiatych a začiatkom osemdesiatych rokov, J. Ross Quinlan, výskumník v oblasti strojového učenia, vyvinul algoritmus rozhodovacieho stromu ID3 (Iterative Dichotomiser), ktorý využíva koncept informačného zisku. Vylepšením a doplnením daného algoritmu bol vytvorený algoritmus označený C4.5. Líši sa vo výbere atribútov, ktoré sa podľa normalizovaného informačného zisku. V roku 1984 zverejnila skupina štatistikov, L. Breiman, J. Friedman, R. Olshen a C. Stone, knihu *Classification and Regression Trees*, v ktorej opísali generáciu binárnych rozhodovacích stromov pomenovaných CART. [31]

Pri vytváraní algoritmu rozhodovacieho stromu v prostredí R nám pomohol balíček *rpart*, ktorý slúži pre výpočet daného algoritmu.

4.2.3. Náhodný les

Algoritmus náhodný les, anglicky *Random forest*, môže byť použitý pre klasifikáciu aj regresiu. Princíp ich fungovania spočíva v generovaní veľkého počtu navzájom dekorelovaných rozhodovacích stromov a následného zoskupenia ich výsledkov. Považujú sa teda za nadstavbu rozhodovacích stromov a odstraňuje ich viaceré nedostatky, napríklad ich nestabilitu. Ďalej, veľmi veľké rozhodovacie stromy, ktoré neupravíme orezaním fungujú výborne na trénovacej množine, avšak na testovacej množine zlyhávajú. Tento problém rieši taktiež náhodný les. Daný algoritmus sa skladá z viacerých rozhodovacích stromov CART, kde je dostupná len náhodná podmnožina atribútov, ktorá je nezávislá od ich počtu. Náhodný les je závislí od sily jednotlivých klasifikátorov a od miery závislosti medzi nimi. [31]

Pri vytváraní daného algoritmu v prostredí R nám pomohol balíček *randomForest*, ktorý slúžil pre výpočet algoritmu náhodného lesa.

4.2.4. Podporné vektory

Podporné vektory, anglicky *Support Vector Machines*, je algoritmus vykonávajúci klasifikáciu, ale aj regresnú analýzu. Dokáže pracovať s dátami, ktoré sú oddelené lineárne ale aj nelineárne. Cieľom daného algoritmu je nájsť tzv. nadrovinu, ktorá rozdeľuje dve triedy a určuje, ktoré body patria do ktorej triedy. Optimálna nadrovina je taká, ktorá je umiestnená v čo najväčšom odstupe od krajných bodov, ktoré sa nazývajú podporné vektory. Jednoduchšie povedané, nadrovina okolo seba vytvára čo najväčšie pásmo bez bodov, s tým, že okrajové body rozdelených oblastí sú rovnako vzdialené od stredu pásma nadroviny. Dôležitá súčasť techniky podporných vektorov je jadrová transformácia priestoru príznakov dát do priestoru transformovaných príznakov spravidla vyššej dimenzie. Táto transformácia nám umožňuje previesť pôvodne lineárne neseparovateľnú úlohu na úlohu lineárne separovateľnú, na ktorú môžeme následne aplikovať optimalizačný algoritmus pre nájdenie rozdeľujúcej nadroviny. Pri výpočte nadroviny sa využíva skalárny súčin transformovaných dát. [31]

Pri vytváraní daného algoritmu v prostredí R nám pomohol balíček *e1071*, ktorý slúži pre jeho výpočet.

4.3. Pochopenie cieľa

Cieľom našej práce z obchodného hľadiska je nájdenie vhodného modelu, ktorý bude, čo najefektívnejšie detekovať falošné správy. Z technického hľadiska je potrebné vytvoriť modely, ktoré následne vyhodnotíme pomocou kontingenčnej tabuľky (confusion matrix), ktorá klasifikuje prvky sledovaného súboru do dvoch skupín podľa pravidiel klasifikácie. Cieľový atribút teda v našom prípade nadobúda hodnoty ***FAKE*** alebo ***REAL***. Pomocou danej matice teda vieme určiť, či bola správa falošná alebo pravdivá.

		Skutočný stav	
		FAKE	**REAL**
Predikcia	**FAKE**	***True positive (TP)*** správne zaradené	***False positive (FP)*** nesprávne zaradené
	REAL	***False negative (FN)*** nesprávne nezaradené	***True negative (TN)*** správne nezaradené

Obr. 5 Kontingenčná tabuľka

[28]

Matica obsahuje skutočné a predikované hodnoty, ktorých kombinácie sú:

- ***true positive*** - správa bola predikovaná ako *„FAKE“* a v skutočnosti bola *„FAKE“*,
- ***false positive*** - správa bola predikovaná ako *„FAKE“* a v skutočnosti bola *„REAL“*,
- ***false negative*** - správa bola predikovaná ako *„REAL“* a v skutočnosti bola *„FAKE“*,
- ***true negative*** - správa bola predikovaná ako *„REAL“* a v skutočnosti bola *„REAL“*.

Model bude najpresnejší, ak docielime nasledujúce pravidlá:

- ***true positive*** a ***true negative*** budú dosahovať čo **najvyššie** hodnoty,
- ***false positive*** a ***falsie negative*** budú dosahovať čo **najnižšie** hodnoty.

Pomocou kontingenčnej tabuľky následne vyhodnotíme vytvorené modely pomocou ukazovateľov binárnej klasifikácie:

- ***spoľahlivosť (accuracy)*** - pomer skutočných výsledkov ku všetkým prípadom [28]:

$$ACC = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN}$$

- ***konfidenčný interval (confidence interval)*** - interval spoľahlivosti, ktorý určuje úroveň spoľahlivosti parametra ležiaceho v danom intervale,

- ***citlivosť, návratnosť (sensitivity, recall)*** - určuje podiel správne zaradených výsledkov ku všetkým pozitívnym, preto sa aj inak nazýva skutočná pozitívna hodnota [28]:

$$\boldsymbol{TPR = \frac{TP}{P} = \frac{TP}{TP + FN}}$$

- ***špecifickosť (specificity)*** - určuje podiel správne nezaradených výsledkov ku všetkým negatívnym, preto sa aj inak označuje ako skutočná záporná hodnota [28]:

$$\boldsymbol{TNR = \frac{TN}{N} = \frac{TN}{TN + FP}}$$

- ***presnosť (precision)*** - pomer správne zaradených výsledkov ku správne a nesprávne zaradeným výsledkom, inak sa označuje ako pozitívna prediktívna hodnota [28]:

$$\boldsymbol{PPV = \frac{TP}{TP + FP}}$$

- ***harmonický priemer (F1 miera)*** - vyjadruje rovnováhu medzi presnosťou (precision) a citlivosťou (sensitivity). Ak skóre dosiahne hodnotu 1, hovoríme o ideálnej správnosti a citlivosti, naopak 0 znamená zlyhanie.

$$\boldsymbol{F_1 = 2 * \frac{PPV * TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}}$$

4.4. Pochopenie dát

Pre riešenie danej práce sme vybrali dve dátové množiny, ktoré sú voľne dostupné na stránke https://www.kaggle.com/.

Prvá dátová množina, *Real or fake*, obsahuje 6335 záznamov a 4 atribúty. Označenie správ, ktoré reprezentuje cieľový atribút je určené pomocou atribútu *label*, ktorý nadobúda dve hodnoty a to *FAKE* - správa je falošná alebo *REAL* - správa je pravdivá. Podiel falošných správ k pravdivým je

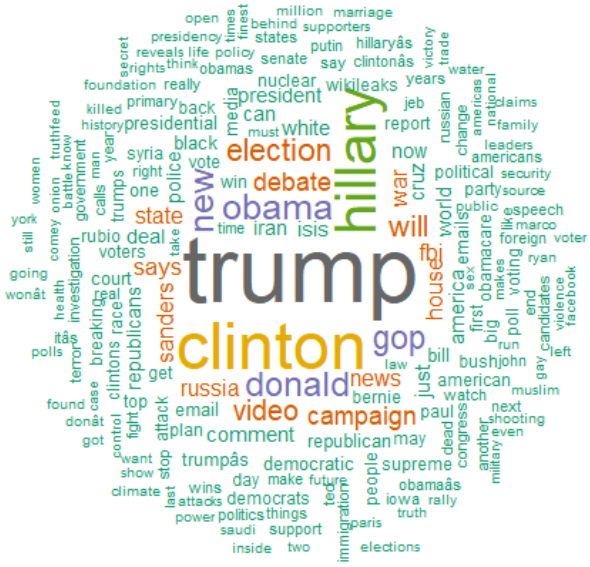
49,94% ku 50,06%. Môžeme vidieť, že rozloženie cieľového atribútu je veľmi rovnomerné. Tabuľka nižšie opisuje charakteristiku atribútov danej dátovej množiny.

Názov atribútu	*Typ atribútu*	*Popis atribútu*
	Numerický	Atribút nemá pomenovanie, obsahuje číselné údaje, identifikačné číslo správy
Title	Znak	Obsahuje názov správy
Text	Znak	Obsahuje telo správy
Label	Znak	Predstavuje označenie správy

Tab. 1 Popis atribútov prvej dátovej množiny

Dátovú množinu vizualizujeme pomocou wordcloudu, ktorý nám zobrazí najčastejšie používané slová do podoby tzv. mraku slov. Wordcloud je najviac používaným nástrojom pre vizualizáciu textových dát, nakoľko je tzv. mrak slov viac priehľadnejší, jednoduchší, zrozumiteľnejší a vizuálne viac pútavý ako údajové tabuľky. Balíky, ktoré sme pri vytváraní použili:

- *tm* - rámec pre analýzu textu,
- *wordcloud* - generátor tzv. mraku slov.



Obr. 6 Vizualizácia prvej dátovej množiny

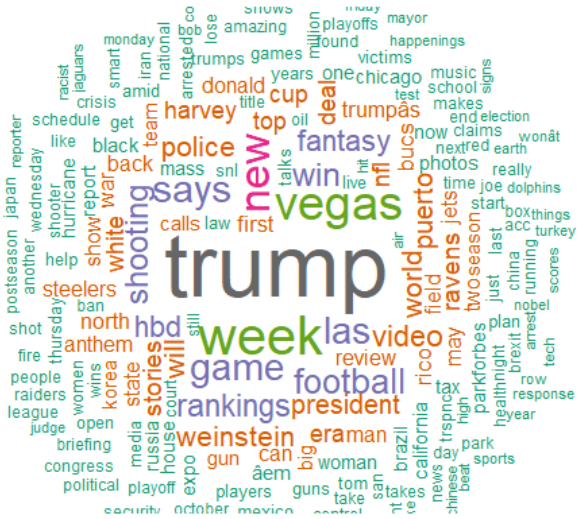
Ako môžeme vidieť, prvá dátová množina obsahuje správy o prezidentských voľbách v USA z roku 2016. Najčastejšie sa vyskytujúcimi slovami sú práve mená, ktoré boli počas prezidentských volieb najviac spomínané, a to, Trump, Clinton, či Obama.

Druhá dátová množina, *Fake News detection*, obsahuje 4009 záznamov a 4 atribúty. Označenie správ je určené pomocou atribútu *Label*, ktorý nadobúda dve hodnoty a to *0* - označuje falošné správy a *1* - označuje pravdivé správy. Daný atribút reprezentuje cieľový atribút. Podiel falošných a pravdivých správ je 51,31% ku 46,69%. Ako pri prvej dátovej množine, aj tu môžeme vidieť, že rozloženie cieľového atribútu je veľmi rovnomerné. Tabuľka nižšie opisuje charakteristiku atribútov danej dátovej množiny.

Názov atribútu	*Typ atribútu*	*Popis atribútu*
URLs	Znak	Označuje umiestnenie správy na internete
Headline	Znak	Obsahuje názov správy
Body	Znak	Obsahuje telo správy
Label	Znak	Predstavuje označenie správy

Tab. 2 Popis atribútov druhej dátovej množiny

Rovnako ako pri prvej dátovej množine aj tu dáta vizualizujeme pomocou wordcloudu, tzv. mraku slov, ktorý nám zobrazí najviac používané slová.



Obr. 7 Vizualizácia druhej dátovej množiny

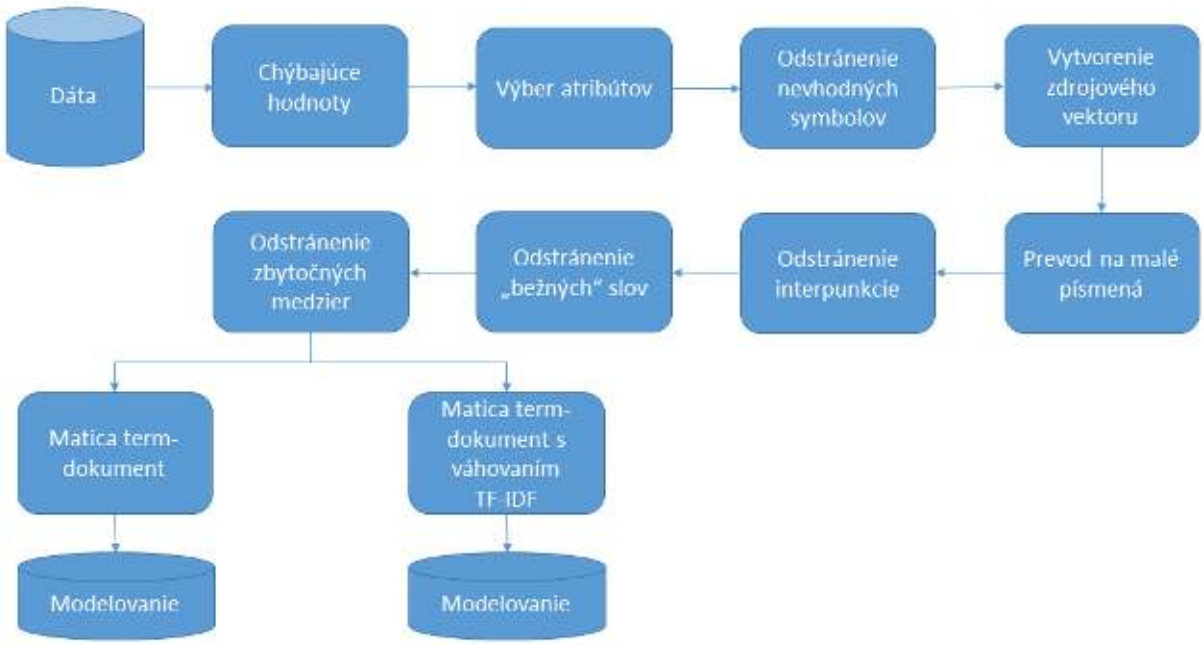
Z vyššie uvedeného obrázku je možné vidieť, že druhá dátová množina obsahuje správy o novom prezidentovi USA Donaldovi Trumpovi. Najčastejšie sa vyskytujúcim slovom je práve jeho meno a ďalšie výrazy, ktoré sú s nim spojené ako new (nový), president (prezident).

4.5. Príprava dát

Príprava dát je jednou z najnáročnejších fáz procesu, nakoľko kvalita vstupných dát vplýva na kvalitu výstupných modelov a tým aj na kvalitu dosiahnutých výsledkov. Obe dátové množiny sme predspracovali na základe nižšie zobrazeného algoritmu. Pri druhej dátovej množine, *Fake News detection*, bolo ešte potrebné pred začatím predspracovanie nahradiť v atribúte *Label* hodnoty:

- 0 --> FAKE,
- 1 --> REAL.

Nasleduje opis každého kroku predspracovania, ktorý bol vykonávaný v prostredí R Studio a teda pomocou jazyka R. Balíček *tm* bol potrebný na vykonanie nasledujúcich krokov úpravy dát.



Obr. 8 Predspracovanie dát

Chýbajúce hodnoty

V oboch dátových množinách sa nenachádzajú chýbajúce hodnoty. Zistili sme to pomocou funkcie *is.na*.

Výber atribútov

Každá dátová množina obsahuje názov správy, telo správy a jej označenie. Každý model, ktorý budeme vytvárať bude najprv pracovať s cieľovým atribútom, ktorým je *label* (označenie správy) a *názvom správy*. Potom bude následne každý model pracovať s atribútom *label* a *telom správy*. Tým nám vzniknú štyri experimenty, ktoré budeme vyhodnocovať, preto je potrebné dbať na výber atribútov. V tabuľke nižšie vizuálne zobrazujeme dané experimenty a atribúty, ktoré budú obsahovať.

Experimenty	*Dátová množina*	*Atribúty*
Experiment č.1	Real or fake	Title, Label
Experiment č.2	Real or fake	Text, Label
Experiment č.3	Fake News detection	Headline, Label
Experiment č.4	Fake News detection	Body, Label

Tab. 3 Rozloženie atribútov v jednotlivých experimentoch

Odstránenie nevhodných symbolov

Pri bližšom skúmaní dátových množín sme našli symboly, ktoré bolo potrebné odstrániť a nahradiť ich prázdnym miestom. Dáta obsahovali symboly *„â€™, â€“*, ktoré sme odstránili pomocou funkcie *str_replace_all*.

Vytvorenie zdrojového vektoru

Pomocou funkcie *VectorSource* sme vytvorili zdrojový vektor, označovaný ako korpus, do ktorého sme vložili textový atribút, s ktorým sme vedeli následne vykonávať spracovanie textu. Textový atribút bol prekonvertovaný na vektor.

Prevod na malé písmená

Pri pracovaní s textom je potrebné zjednotenie písmen. Pomocou funkcie *tm_map* pri zadefinovaní argumentu *tolower* boli zmenené všetky znaky na malé písmená.

Odstránenie interpunkcie

Vo všeobecnosti interpunkcia nepridáva žiadnu hodnotu pri analýze textu s využitím modelov klasifikácie. Z daného dôvodu je potrebné interpunkčné znamienka z dátových množín vymazať za pomoci funkcie *tm_map* zadefinovaním argumentu *removePunctuation.*

Odstránenie „bežných" slov

Ďalšou fázou predspracovania textových dát je odstránenie tzv. „bežných" slov, anglický výraz „stopwords". Sú to slová bez významu, ktoré sa používajú bežne vo vetách pri ich spájaní a ich informačná hodnota pri analýze textu je nulová. Členy, predložky, spojky a niektoré zámená sa považujú za tieto tzv. stopwords. Je možné vytvorenie vlastného zoznamu týchto slov, ktoré chceme z dátovej množiny odstrániť. Pri riešení reálnych úloh sa však využívajú typické slová konkrétneho jazyka. [29] Naše dáta sú anglickými správami, preto sme pri odstraňovaní „bežných" slov použili funkciu *tm_map* pri ktorej sme zadefinovali argument *stopwords('english')*.

Odstránenie zbytočných medzier

Predspracovanie, ktoré sme doteraz vykonali, nám v dokumente ponechalo množstvo medzier, ktoré neboli odstránené spolu so slovami, ktoré boli vyššie uvedeným predspracovaním vymazané. Častokrát sú tieto medzery považované za slovo, preto je potrebné tieto biele miesta vymazať pomocou funkcie *tm_map* zadefinovaním argumentu *stripWhitespace*.

Matica term-dokument

Matica term-dokument, anglicky document term matrix (DTM), je najbežnejším spôsobom reprezentácie textu pre ďalšie spracovanie. V našom prípade exportujeme predspracované dáta do podoby tzv. „vreca slov", anglicky „bag of words". Tento model je irelevantný, čo znamená, že nezohľadňuje poradie slov v dokumente. Matica opisuje frekvenciu slov, ktoré sa v súbore dokumentu vyskytujú. Riadky matice predstavujú vybrané termy a stĺpce jednotlivé dokumenty. [29] V jazyku R danú maticu vytvoríme veľmi jednoducho a to použitím funkcie *DocumentTermMatrix*.

Matica term-dokument s váhovaním

Ďalšou maticou, s ktorou budeme pracovať bude matica term-dokument s váhovaním TF-IDF. Momentálne je v dnešnej dobe najpopulárnejšia váhovacia schéma TF-IDF, dokumentová frekvencia - inverzná dokumentová frekvencia, anglicky term frequency - inverse document frequency, ktorej cieľom je vyjadriť dôležitosť slova pre dokument v korpuse. Priamoúmerne sa zvyšuje s počtom slov objaveným v dokumente ale je kompenzovaný svojou frekvenciou v celkovom korpuse. Daná metrika neberie do úvahy pozíciu alebo kontext slova. Je vyjadrená súčinom váh *tf* a *idf*: [30]

$$\mathbf{w_{i,j}} = \mathbf{tf_{i,j}} * \mathbf{log}(\frac{\mathbf{N}}{\mathbf{df_i}})$$

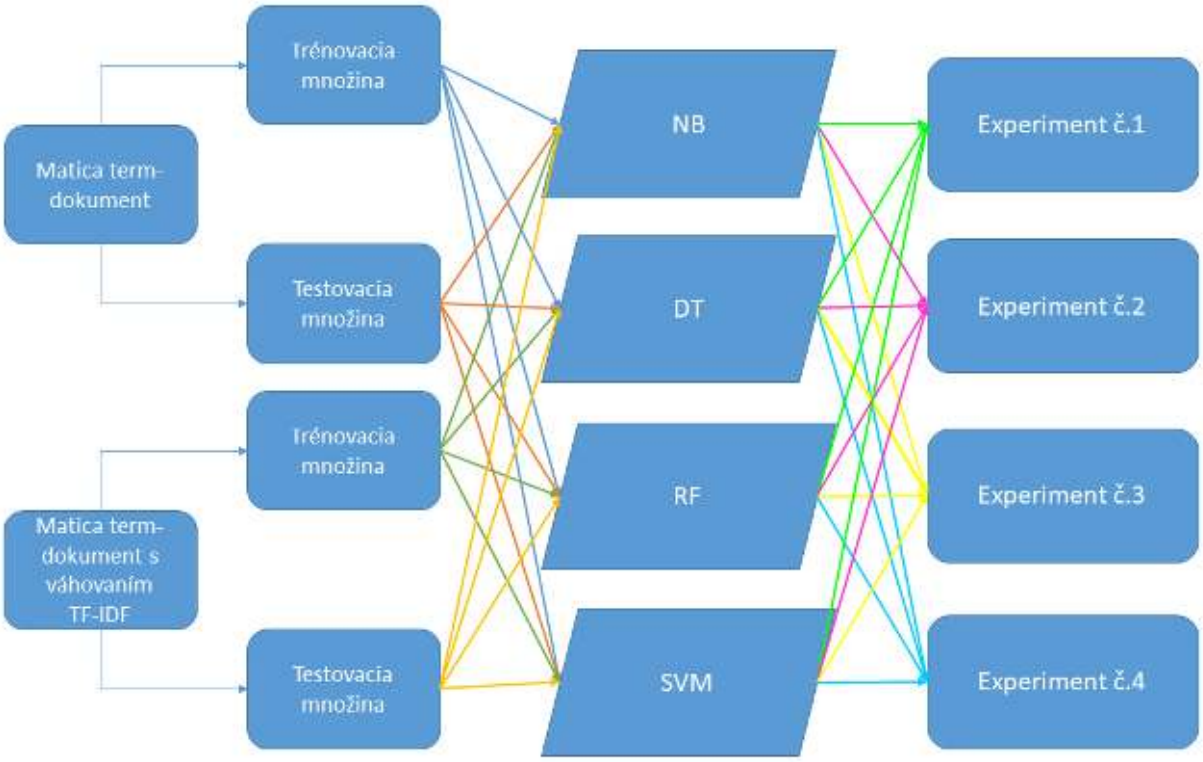
Maticu term-dokument s váhovaním TF-IDF vytvoríme v jazyku R funkciou *DocumentTermMatrix* pridaním argumentu *weighting*, váha, *weightTfIdf*.

4.6. Modelovanie

Fáza modelovania je kľúčovou pre dosiahnutie hľadaných výsledkov. Pri modelovaní sme na trénovaciu množinu aplikovali algoritmy strojového učenia, konkrétne *Naive Bayes, rozhodovacie stromy, náhodný les* a *podporné vektory.* Následne sme modely overovali na testovacej množine. Pracovali sme v prostredí R Studio pomocou jazyka R. Dátové množiny sme na začiatku rozdelili v pomere 70% trénovacia množina a 30% testovacia množina. Ako sme uviedli v kapitole 3.2., modely sme vyhodnocovali pomocou ukazovateľov binárnej klasifikácie, kde sme konkrétne sledovali:

- spoľahlivosť a interval spoľahlivosti,
- citlivosť a špecifickosť,
- návratnosť, presnosť a F1 miera.

V kapitole 3.4. sme zobrazili štyri experimenty, na ktoré budú aplikované vybrané algoritmy. Na záver ich vyhodnotíme pomocou ukazovateľov, ktoré sme vyššie rozdelili do troch skupín. Z toho vyplýva, že sledovať budeme štyri experimenty, ktoré budú obsahovať tri tabuľky výsledkov. Zároveň budeme sledovať výsledky pri zmene vstupnej reprezentácie textu, a to ako matica term-dokument a matica term-dokument s váhovaním TF-IDF. Nižšie vizuálne zobrazíme postup fázy modelovania. Ďalej opíšeme a vizuálne zobrazíme experimenty.



Obr. 9 Modelovanie

4.6.1. Experiment č.1

V experimente č. 1 sme pracovali s prvou dátovou množinou, *Real or fake,* a s atribútom *title*, ktorý nám uvádza nadpis príspevku. Daný atribút obsahuje pomerne menšie množstvo slov ako atribút *text*, s ktorým budeme pracovať v experimente č.2. Reprezentácia vstupov je vo forme matica term-dokument a matica term-dokument s váhovaním TF-IDF. Na dané vstupy sme aplikovali algoritmy *Naive Bayes, Decision Tree, Random Forest a Support Vector Machines.* Sledovanými ukazovateľmi boli *spoľahlivosť a interval spoľahlivosti, citlivosť a špecifickosť, v závere návratnosť, presnosť a F1 miera.* V tabuľkách nižšie sú dané ukazovatele znázornené numericky.

Dátová množina 1 - atribút title	*DTM*		*DTM s váhovaním TF-IDF*	
	Spoľahlivosť	*Interval spoľahlivosti*	*Spoľahlivosť*	*Interval spoľahlivosti*
NB	***0.793***	(0.774, 0.812)	***0.788***	(0.768, 0.806)
DT	0.670	(0.649, 0.691)	0.700	(0.679, 0.720)
RF	0.758	(0.738, 0.777)	0.760	(0.740, 0.779)
SVM	0.767	(0.748, 0.786)	0.776	(0.756, 0.794)

Tab. 4 Experiment č.1 - sledovanie presnosti a intervalu spoľahlivosti

Pri sledovaní presnosti a intervalu spoľahlivosti je možné vidieť, že najvyššiu spoľahlivosť dosiahol model ***Naive Bayes*** pri matici term-dokument. Ostatné modely dosiahli nižšiu spoľahlivosť, no pri každom modely je jasne vidieť, že vyššia spoľahlivosť bola dosiahnutá pri matici term-dokument s váhovaním TF-IDF.

Dátová množina 1 - atribút title	*DTM*		*DTM s váhovaním TF-IDF*	
	Citlivosť	*Špecifickosť*	*Citlivosť*	*Špecifickosť*
NB	***0.785***	***0.802***	***0.781***	***0.794***
DT	0.931	0.399	0.844	0.549
RF	0.734	0.782	0.720	0.802
SVM	0.794	0.741	0.777	0.775

Tab. 5 Experiment č.1 - sledovanie citlivosti a špecifickosti

Citlivosť nám určuje podiel správne zaradených výsledkov ku všetkým pozitívnym na rozdiel od špecifickosti, ktorý určuje podiel správne zaradených negatívnych výsledkov ku všetkým negatívnym. Aj pri sledovaní týchto ukazovateľov je zrejmé, že najlepšie výsledky dosiahol model ***Naive Bayes*** ako pri predchádzajúcom sledovaní. Avšak tu je viditeľný fakt, že pri modely *Decision*

Tree boli nesprávne predikované správy, konkrétne bol vysoký počet správ označených ako *„FAKE“* a v skutočnosti boli *„REAL“.*

Dátová množina 1 - atribút title	*DTM*			*DTM s váhovaním TF-IDF*		
	Návratnosť	*Presnosť*	*F1 miera*	*Návratnosť*	*Presnosť*	*F1 miera*
NB	***0.785***	***0.804***	***0.794***	***0.781***	***0.797***	***0.789***
DT	0.931	0.617	0.742	0.844	0.661	0.741
RF	0.734	0.778	0.755	0.720	0.791	0.754
SVM	0.794	0.754	0.773	0.777	0.775	0.776

Tab. 6 Experiment č.1 - sledovanie citlivosti, správnosti a harmonického priemeru

Pri sledovaní ukazovateľov tretej skupiny dosiahol najlepšie hodnoty taktiež model ***Naive Bayes***, ako pri predošlých pozorovaniach. Harmonický priemer alebo F1 miera vyjadruje rovnováhu medzi presnosťou (precision) a citlivosťou (sensitivity). Hodnota sa približuje k hodnote 1, čo nám poukazuje na fakt ideálnej správnosti a citlivosti.

4.6.2. Experiment č.2

Experiment č. 2 pracuje s prvou dátovou množinou, *Real or fake,* a atribútom *text*, ktorý nám uvádza text príspevku. Daný atribút obsahuje väčšie množstvo slov ako atribút *title*, s ktorým sme pracovali v experimente č.1. Reprezentácia vstupov je vo forme matica term-dokument a matica term-dokument s váhovaním TF-IDF. Na dané vstupy sme aplikovali algoritmy *Naive Bayes, Decision Tree, Random Forest a Support Vector Machines.* Sledovanými ukazovateľmi boli *spoľahlivosť a interval spoľahlivosti, citlivosť a špecifickosť, v závere návratnosť, presnosť a F1 miera.* V tabuľkách nižšie sú dané ukazovatele numericky znázornené.

Dátová množina 1 - atribút text	*DTM*		*DTM s váhovaním TF-IDF*	
	Spoľahlivosť	*Interval spoľahlivosti*	*Spoľahlivosť*	*Interval spoľahlivosti*
NB	0.813	(0.795, 0.831)	0.747	(0.727, 0.767)
DT	0.768	(0.748, 0.787)	0.778	(0.758, 0.796)
RF	***0.900***	(0.885, 0.913)	***0.906***	(0.892, 0.919)
SVM	0.713	(0.682, 0.742)	0.898	(0.883, 0.911)

Tab. 7 Experiment č.2 - sledovanie presnosti a intervalu spoľahlivosti

Pri sledovaní presnosti a intervalu spoľahlivosti je možné vidieť, že najvyššiu spoľahlivosť dosiahol model ***Random Forest*** pri matici term-dokument s váhovaním TF-IDF. Ostatné modely dosiahli nižšiu spoľahlivosť kde je možné vidieť, že okrem modelu *Naive Bayes*, dosiahol každý model vyššiu spoľahlivosť pri matici term-dokument s váhovaním TF-IDF. Model *Naive Bayes* dosiahol ako pri experimente č.1 nižšiu spoľahlivosť pri matici term-dokument, ako keď bol vstup reprezentovaný maticou term-dokument s váhovaním TF-IDF.

Dátová množina 1 - atribút text	*DTM*		*DTM s váhovaním TF-IDF*	
	Citlivosť	*Špecifickosť*	*Citlivosť*	*Špecifickosť*
NB	0.886	0.738	0.789	0.704
DT	0.792	0.744	0.822	0.732
RF	***0.895***	***0.905***	***0.902***	***0.911***
SVM	0.475	0.947	0.913	0.883

Tab. 8 Experiment č.2 - sledovanie citlivosti a špecifickosti

Pri sledovaní ukazovateľov citlivosti a špecifickosti je zrejmé, že najlepšie výsledky dosiahol model ***Random Forest*** ako pri predchádzajúcom sledovaní. Avšak tu je viditeľný fakt, že pri modely *Support Vector Machine* pri matici term-dokument boli nesprávne predikované správy, konkrétne bol vysoký počet správ označených ako *„REAL“* a v skutočnosti boli *„FAKE“.*

Dátová množina 1 - atribút text	*DTM*			*DTM s váhovaním TF-IDF*		
	Návratnosť	*Presnosť*	*F1 miera*	*Návratnosť*	*Presnosť*	*F1 miera*
NB	0.886	0.777	0.828	0.789	0.733	0.760
DT	0.792	0.762	0.777	0.822	0.761	0.790
RF	***0.895***	***0.907***	***0.901***	***0.902***	***0.913***	***0.907***
SVM	0.475	0.898	0.621	0.913	0.886	0.899

Tab. 9 Experiment č.2 - sledovanie citlivosti, správnosti a harmonického priemeru

Pri sledovaní ukazovateľov tretej skupiny dosiahol najlepšie hodnoty taktiež model ***Random Forest***, ako pri predošlých pozorovaniach. Harmonický priemer alebo F1 miera vyjadruje rovnováhu medzi presnosťou (precision) a citlivosťou (sensitivity). Hodnota je veľmi blízko hodnoty 1, čo nám poukazuje na fakt ideálnej správnosti a citlivosti.

4.6.3. Experiment č.3

V experimente č. 3 sme pracovali s druhou dátovou množinou, *Fake News detection,* a s atribútom *headline*, ktorý nám uvádza nadpis príspevku. Daný atribút obsahuje pomerne menšie množstvo slov ako atribút *body*, s ktorým budeme pracovať v experimente č.4. Reprezentácia vstupov je vo forme matica term-dokument a matica term-dokument s váhovaním TF-IDF. Na dané vstupy sme aplikovali algoritmy *Naive Bayes, Decision Tree, Random Forest a Support Vector Machines.* Sledovanými ukazovateľmi boli *spoľahlivosť a interval spoľahlivosti, citlivosť a špecifickosť, v závere návratnosť, presnosť a F1 miera.* V tabuľkách nižšie sú dané ukazovatele znázornené numericky.

Dátová množina 2 - atribút headline	*DTM*		*DTM s váhovaním TF-IDF*	
	Spoľahlivosť	*Interval spoľahlivosti*	*Spoľahlivosť*	*Interval spoľahlivosti*
NB	***0.802***	(0.778, 0.824)	***0.812***	(0.788, 0.833)
DT	0.551	(0.523, 0.580)	0.550	(0.521, 0.578)
RF	0.749	(0.724, 0.773)	0.760	(0.735, 0.784)
SVM	0.762	(0.737, 0.786)	0.775	(0.750, 0.798)

Tab. 10 Experiment č.3 - sledovanie presnosti a intervalu spoľahlivosti

Pri sledovaní presnosti a intervalu spoľahlivosti je možné vidieť, že najvyššiu spoľahlivosť dosiahol model ***Naive Bayes*** pri matici term-dokument s váhovaním TF-IDF. Ostatné modely dosiahli nižšiu spoľahlivosť, kde je možné vidieť, že okrem modelu *Decision Tree*, dosiahol každý model vyššiu spoľahlivosť pri matici term-dokument s váhovaním TF-IDF. Pri modely *Decision Tree* je rozdiel len 0.001.

Dátová množina 2 - atribút headline	*DTM*		*DTM s váhovaním TF-IDF*	
	Citlivosť	*Špecifickosť*	*Citlivosť*	*Špecifickosť*
NB	***0.792***	***0.813***	***0.820***	***0.802***
DT	1.000	0.040	0.200	0.972
RF	0.830	0.657	0.710	0.820
SVM	0.718	0.813	0.766	0.784

Tab. 11 Experiment č.3 - sledovanie citlivosti a špecifickosti

Pri sledovaní ukazovateľov citlivosti a špecifickosti je zrejmé, že najlepšie výsledky dosiahol model ***Naive Bayes*** ako pri predchádzajúcom sledovaní. Avšak tu je viditeľný fakt, že pri modely *Decision Tree* boli nesprávne predikované správy, konkrétne bol vysoký počet správ označených ako *„FAKE“* a v skutočnosti boli *„REAL“* pri matici term-dokument. Na druhej strane, keď bol vstup reprezentovaný maticou term-dokument s váhovaním TF-IDF, bol vysoký počet správ predikovaných ako *„REAL“* a v skutočnosti boli *„FAKE“*.

Dátová množina 2 - atribút headline	*DTM*			*DTM s váhovaním TF-IDF*		
	Návratnosť	*Presnosť*	*F1 miera*	*Návratnosť*	*Presnosť*	*F1 miera*
NB	***0.792***	***0.829***	***0.810***	***0.820***	***0.826***	***0.823***
DT	1.00	0.543	0.704	0.200	0.898	0.327
RF	0.830	0.734	0.779	0.710	0.827	0.764
SVM	0.718	0.814	0.763	0.766	0.802	0.783

Tab. 12 Experiment č.3 - sledovanie citlivosti, správnosti a harmonického priemeru

Pri sledovaní ukazovateľov tretej skupiny dosiahol najlepšie hodnoty taktiež model ***Naive Bayes***, ako pri predošlých pozorovaniach. Harmonický priemer alebo F1 miera vyjadruje rovnováhu medzi presnosťou (precision) a citlivosťou (sensitivity). Hodnota sa približuje k hodnote 1, čo nám poukazuje na fakt ideálnej správnosti a citlivosti. F1 miera pri modely *Decision Tree* pri matici term-dokument s váhovaním TF-IDF je bližšie k hodnote 0 ako k 1, čo nám potvrdzuje fakt, ktorý sme opísali vyššie, a to, že hodnoty v kontingenčnej tabuľke, konkrétne *false positive* a *false negative* dosahovali veľmi vysoké hodnoty.

4.6.4. Experiment č.4

Experiment č. 4 pracuje s druhou dátovou množinou, *Fake News detection,* a atribútom *body*, ktorý nám uvádza text príspevku. Daný atribút obsahuje väčšie množstvo slov ako atribút *headline*, s ktorým sme pracovali v experimente č.3. Reprezentácia vstupov je vo forme matica term-dokument a matica term-dokument s váhovaním TF-IDF. Na dané vstupy sme aplikovali algoritmy *Naive Bayes, Decision Tree, Random Forest a Support Vector Machines.* Sledovanými

ukazovateľmi boli *spoľahlivosť a interval spoľahlivosti, citlivosť a špecifickosť, v závere návratnosť, presnosť a F1 miera.* V tabuľkách nižšie sú dané ukazovatele numericky znázornené.

Dátová množina 2 - atribút body	*DTM*		*DTM s váhovaním TF-IDF*	
	Spoľahlivosť	*Interval spoľahlivosti*	*Spoľahlivosť*	*Interval spoľahlivosti*
NB	0.844	(0.822, 0.864)	0.904	(0.886, 0.920)
DT	0.881	(0.862, 0.899)	0.904	(0.886, 0.920)
RF	***0.978***	(0.969, 0.988)	***0.983***	(0.973, 0.989)
SVM	0.782	(0.758, 0.805)	0.944	(0.930, 0.957)

Tab. 13 Experiment č.4 - sledovanie presnosti a intervalu spoľahlivosti

Pri sledovaní presnosti a intervalu spoľahlivosti je možné vidieť, že najvyššiu spoľahlivosť dosiahol model ***Random Forest*** pri matici term-dokument s váhovaním TF-IDF. Ostatné modely dosiahli nižšiu spoľahlivosť ale každý model dosiahol vyššia spoľahlivosť pri matici term-dokument s váhovaním TF-IDF.

Dátová množina 2 - atribút body	*DTM*		*DTM s váhovaním TF-IDF*	
	Citlivosť	*Špecifickosť*	*Citlivosť*	*Špecifickosť*
NB	0.910	0.768	0.880	0.930
DT	0.938	0.817	0.920	0.886
RF	***0.964***	***0.995***	***0.972***	***0.995***
SVM	0.570	0.967	0.916	0.977

Tab. 14 Experiment č.4 - sledovanie citlivosti a špecifickosti

Pri sledovaní ukazovateľov citlivosti a špecifickosti je zrejmé, že najlepšie výsledky dosiahol model ***Random Forest*** ako pri predchádzajúcom sledovaní. Hodnota citlivosti sa nám od ostatných líši pri modely *Support Vector Machine* pri matici term-dokument, kde boli nesprávne predikované správy, konkrétne bol vysoký počet správ označených ako *„REAL“* a v skutočnosti boli *„FAKE“.*

Dátová množina 2 - atribút body	*DTM*			*DTM s váhovaním TF-IDF*		
	Návratnosť	*Presnosť*	*F1 miera*	*Návratnosť*	*Presnosť*	*F1 miera*
NB	0.910	0.818	0.862	0.880	0.936	0.907
DT	0.938	0.854	0.894	0.920	0.902	0.911
RF	***0.964***	***0.995***	***0.979***	***0.972***	***0.995***	***0.983***
SVM	0.570	0.938	0.709	0.916	0.978	0.946

Tab. 15 Experiment č.4 - sledovanie citlivosti, správnosti a harmonického priemeru

Pri sledovaní ukazovateľov tretej skupiny dosiahol najlepšie hodnoty taktiež model ***Random Forest***, ako pri predošlých pozorovaniach. Harmonický priemer alebo F1 miera vyjadruje rovnováhu medzi presnosťou (precision) a citlivosťou (sensitivity). Hodnota siaha k hodnote 1, čo nám poukazuje na fakt ideálnej správnosti a citlivosti.

4.7. Vyhodnotenie výsledkov

Na záver zhrnieme experimenty a popíšeme vyhodnotenia, ktoré nám dané experimenty priniesli:

- Experiment č.4 hodnotíme za najlepší z pomedzi všetkých štyroch experimentov, nakoľko dosahoval najvyššie výsledky pri všetkých model.
- Všetky experimenty nám dokazujú, že vstupná reprezentácia dát je veľmi dôležitá, nakoľko lepšie výsledky dosahovali vo väčšine prípadov modely, ktoré pracovali s maticou term-dokument s váhovaním TF-IDF.
- V experimente č.1 a č.3 dosahoval najlepšie výsledky model Naive Bayes, kde vstupmi bola menšia dátová množina, nakoľko sa v týchto prípadoch jednalo o nadpisy

príspevkov. Môžeme teda konštatovať, že v našom prípade sa javil model Naive Bayes najlepší pri menšej dátovej vstupnej vzorke.

- V experimente č.2 a č.4 dosahoval najlepšie výsledky model Random Forest. Pri týchto experimentoch bola na vstupe väčšia dátová množina, ktorá pozostávala z textu príspevkov. Môžeme v tomto prípade konštatovať, že v našom prípade sa javil model Random Forest za najlepší pri väčšej dátovej vstupnej vzorke.
- Najhoršími modelmi sa v našom prípade javili modely Support Vector Machine a Decision Tree, čo malo za následok nesprávne predikované správy.

Záver

Cieľom predloženej diplomovej práce bolo pomocou viacerých experimentov nájsť vhodný model, ktorý by vedel čo najefektívnejšie a najpresnejšie detekovať falošné správy.

Pri riešení úlohy sme postupovali pomocou metodológie CRISP-DM, ktorá predstavuje proces dolovania dát. K dispozícií sme mali dve dátové množiny, ktoré obsahovali názov príspevku, text príspevku a označenie príspevkov, ktoré sa delili na falošné a pravdivé. Cieľovým a sledovaným atribútom bolo práve označenie príspevkov. Atribúty názov a text príspevkov sme upravili v časti prípravy dát pomocou predspracovania textu, ktorý je veľmi dôležitý pri textových dátach. Hlavnou časťou bola fáza modelovania, ktorá pozostávala zo štyroch experimentov. Každý experiment sa líšil vstupnými dátami, kde sme menili atribút a dátovú množinu, z ktorej sme tieto atribúty vyberali. Reprezentácia vstupov bola v dvoch formách a to, *matica term-dokument a matica term-dokument s váhovaním TF-IDF*. Vstupné dáta sme rozdelili na trénovaciu a testovaciu množinu, kde sme na trénovaciu množinu aplikovali modely *Naive Bayes, Decision Tree, Random Forest a Support Vector Machine*. Vytvorené modely sme overovali na testovacej množine, ktoré sme na záver vyhodnotili pomocou viacerých ukazovateľov binárnej klasifikácie. Medzi najdôležitejšie ukazovatele patrili *spoľahlivosť, návratnosť, presnosť a F1 miera.*

Z experimentov sme poukázali na niekoľko faktov, a to, že je veľmi dôležitá reprezentácia vstupov, nakoľko lepšie výsledky boli dosiahnuté pri matici term-dokument s váhovaním TF-IDF. Ďalej, pri menšom množstve dát dosiahol najlepšie výsledky model Naive Bayes, na druhej strane pri väčšej dátovej množine to bol model Random Forest.

Veríme, že táto diplomová práca priniesla výsledky, s ktorými bude možné ďalej pracovať a rozvíjať ich, nakoľko je téma falošných správ čoraz viac riešená v rámci celého sveta. Bohužiaľ v dobe internetu sa dostávajú na povrch novodobé problémy, pred ktorými by sme nemali zatvárať oči. Práve naopak, o týchto problémoch by sa malo diskutovať, poukazovať na ich nebezpečenstvá a riešiť ich práve nájdením a detekciou.

Zoznam použitej literatúry

[1]. CANN, Alan; DIMITRIOU, Konstantian; HOOLEY, Tristram: *Social Media: A guide for researchers* [online]. [cit. 2018-07-07]. Dostupné z: https://www.researchgate.net/publication/261990960_Social_Media_A_Guide_for_Researchers

[2]. COHEN-ALMAGOR, Raphael: Internet History [online]. [cit. 2018-07-07]. Dostupné z: https://www.researchgate.net/publication/215660523_Internet_History

[3]. AKRAM, W.; KUMAR, R. : *A Study on Positive and Negative Effects of Social Media on Society* [online]. [cit. 2018-07-08]. Dostupné z: https://www.researchgate.net/publication/323903323_A_Study_on_Positive_and_Negative_Effects_of_Social_Media_on_Society

[4]. MARCH, Evita: *'Don't feed the trolls' really is good advice – here's the evidence* [online]. [cit. 2018-07-21]. Dostupné z: https://theconversation.com/dont-feed-the-trolls-really-is-good-advice-heres-the-evidence-63657

[5]. *What is trolling?* [online]. [cit. 2018-07-21]. Dostupné z: https://edu.gcfglobal.org/en/thenow/what-is-trolling/1/

[6]. MOREAU, Elise: *Internet Trolling: How Do You Spot a Real Troll?* [online]. [cit. 2018-07-21]. Dostupné z: https://www.lifewire.com/what-is-internet-trolling-3485891

[7]. WALTERS, Kendall: *How to Deal with Trolls on Social Media* [online]. [cit. 2018-07-21]. Dostupné z: https://blog.hootsuite.com/how-to-deal-with-trolls-on-social-media/

[8]. MOREAU, Elise: *10 Types of Internet Trolls You´ll Meet Online* [online]. [cit. 2018-07-28]. Dostupné z: https://www.lifewire.com/types-of-internet-trolls-3485894

[9]. BRANDON, John: *5 ways to handle comment trolls on social media* [online]. [cit. 2018-07-29]. Dostupné z: https://www.cio.com/article/2935933/online-reputation-management/5-ways-to-handle-comment-trolls-on-social-media.html

[10]. RAMPTON, John: *10 Tips to Dealing With Trolls* [online]. [cit. 2018-07-29]. Dostupné z: https://www.forbes.com/sites/johnrampton/2015/04/09/10-tips-to-dealing-with-trolls/#7348604c54f4

[11]. *Fake news* [online]. [cit. 2018-08-08]. Dostupné z: https://en.wikipedia.org/wiki/Fake_news

[12]. VÍTEK, Filip: *FAKE NEWS – KDE TO ZAČALO A KAM SPEJEME* [online]. [cit. 2018-08-08]. Dostupné z: http://mocnedata.sk/2018-fake-news/

[13]. YAR, Lucia: *Falošné správy sa na internete šíria omnoho rýchlejšie ako pravdivé, tvrdia vedci* [online]. [cit. 2018-08-09]. Dostupné z: https://euractiv.sk/section/digitalizacia/news/falosne-spravy-sa-na-internete-siria-omnoho-rychlejsie-ako-pravdive-tvrdia-vedci/

[14]. SHU, Kai; SLIVA, Amy; WANG, Suhang; TANG, Jiliang; LIU, Huan: *Fake News Detection on Social Media: A Data Mining* [online]. [cit. 2018-08-13].

[15]. *IS ARTIFICIAL INTELLIGENCE THE KEY TO COMBAT FAKE NEWS?* [online]. [cit. 2018-08-14]. Dostupné z: https://www.marutitech.com/artificial-intelligence-fake-news/

[16]. ONG, Thuy: *Facebook found a better way to fight fake news* [online]. [cit. 2018-08-16]. Dostupné z: https://www.theverge.com/2017/12/21/16804912/facebook-disputed-flags-misinformation-newsfeed-fake-news

[17]. VINCENT, James: *Why AI isn´t going to solve Facebook´s fake news problem* [online]. [cit. 2018-08-16]. Dostupné z: https://www.theverge.com/2018/4/5/17202886/facebook-fake-news-moderation-ai-challenges

[18]. *Facebook is using AI to remove fake news* [online]. [cit. 2018-08-16]. Dostupné z: https://www.clickatell.com/articles/digital-marketing/facebook-using-ai-remove-fake-news/

[19]. HREHUŠ, Marián: *Komentár 52. týždeň: Prečo potrebujeme strojové rozpoznávanie falošných správ?* [online]. [cit. 2018-08-16]. Dostupné z: https://techbox.dennikn.sk/temy/komentar-52-tyzden-preco-potrebujeme-strojove-rozpoznavanie-falosnych-sprav/

[20]. MARR, Bernard: *Fake News: How Big Data And AI Can Help* [online]. [cit. 2018-08-20]. Dostupné z: https://www.forbes.com/sites/bernardmarr/2017/03/01/fake-news-how-big-data-and-ai-can-help/#33ce244d70d5

[21]. ROGERS, Kaleigh: *A Data Scientist Was Sick of Seeing Spam on His Facebook so He Built a Fake News Detector* [online]. [cit. 2018-08-20]. Dostupné z: https://motherboard.vice.com/en_us/article/7x7mda/fake-news-detector-neural-network-ai-machine-learning

[22]. SUSARLA, Anjana: *How artificial intelligence can detect – and create – fake news* [online]. [cit. 2018-08-21]. Dostupné z: https://theconversation.com/how-artificial-intelligence-can-detect-and-create-fake-news-95404

[23]. DEMATIS, Ioannis; KARAPISTOLI, Eirini; VAKALI, Athena: *Fake Review Detection via Exploitation of Spam Indicators and Reviewer Behavior Characteristics* [cit. 2018-08-30].

[24] HADEER, Ahmed; ISSA, Traore; SHERIF, Saad: *Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques* [cit. 2019-03-23].

[25] GENES, Yunus: *Detection Fake News With NLP* [cit. 2019-03-23]. Dostupné z: https://medium.com/@Genyunus/detecting-fake-news-with-nlp-c893ec31dee8

[26] CHOWDHARY, Neha S.; PANDIT, Anala A.: *Fake Review Detection using Classification* [cit. 2019-03-23].

[27] L.OLSON, David; DELEN, Dursun: *Advanced Data Mining Techniques,* ISBN: 978-3-540-76916-3 [cit. 2019-03-29].

[28] *Sensitivity and specificity* [online]. [cit. 2019-04-02]. Dostupné z: https://en.wikipedia.org/wiki/Sensitivity_and_specificity

[29] FEINERER, Ingo; HORNIK, Kurt, MEYER, David: *Text Mining Infrastructure in R* [cit. 2019-04-03].

[30] PARALIČ, Ján; FURDÍK, Karol, TUTOKY, Gabriel; BEDNÁR, Peter; SARNOVSKÝ, Martin; BUTKA, Peter; BABIČ, František: *Dolovanie znalostí z textov:* Equilibria: Košice,2010. 188 s. ISBN 978-80-89284-62-7 [cit. 2019-04-03].

[31] HAN, Jiawei; KAMBER, Micheline: *Data Mining: Concepts and Techniques*: Morgan Kaufmann Publisher: San Francisco,2003. 740 s. ISBN 1-55860-901-6 7 [cit. 2019-04-04].

Prílohy

Príloha A: CD médium – diplomová práca v elektronickej podobe, prílohy v elektronickej podobe. CD je spravidla grafické s logom univerzity a fakulty. Pozri . Tieto CD robia v Univerzitnej knižnici TUKE.

Príloha B: Používateľská príručka

Príloha C: Systémová príručka