SLOVENSKÁ TECHNICKÁ
UNIVERZITA V BRATISLAVE
FAKULTA INFORMATIKY
A INFORMAČNÝCH TECHNOLÓGIÍ

Igor Stupavský

Auto referát dizertačnej práce

Antisociálne správanie v online komunitách

na získanie akademického titulu *Philosophiae Doctor (PhD.)*

Študijný program:	Aplikovaná informatika
Študijný odbor:	9.2.9 Aplikovaná informatika
Forma štúdia:	Denná
Ústav:	Ústav informatiky, informačných systémov a softvérového inžinierstva

Bratislava, 2024

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLÓGIÍ

Igor Stupavský

Auto referát dizertačnej práce

Antisociálne správanie v online komunitách

na získanie akademického titulu *Philosophiae Doctor (PhD.)*

Študijný program:	Aplikovaná informatika
Študijný odbor:	9.2.9 Aplikovaná informatika
Forma štúdia:	Denná
Ústav:	Ústav informatiky, informačných systémov a softvérového inžinierstva

Bratislava, 2024

SLOVAK UNIVERSITY OF
TECHNOLOGY IN BRATISLAVA
FACULTY OF INFORMATICS
AND INFORMATION TECHNOLOGIES

Igor Stupavský

Dissertation Thesis Abstract

Antisocial behavior in online communities

to obtain the Academic Title of *Philosophiae Doctor (PhD.)*

Degree course:	Applied Informatics
Field of study:	9.2.9 Aplikovaná informatika
Form of study:	Internal
Place of development:	Institute of Informatics, Information Systems and Software Engineering

Bratislava, 2024

SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA
FACULTY OF INFORMATICS AND INFORMATION TECHNOLOGIES

Igor Stupavský

Dissertation Thesis Abstract

Antisocial behavior in online communities

to obtain the Academic Title of *Philosophiae Doctor (PhD.)*

Študijný program:	Applied Informatics
Študijný odbor:	9.2.9 Aplikovaná informatika
Forma štúdia:	Internal
Ústav:	Institute of Informatics, Information Systems and Software Engineering

Bratislava, 2024

STU

Dizertačná práca bola vypracovaná na/Dissertation Thesis has been prepared at

Ústav informatiky, informačných systémov a softvérového inžinierstva,
Fakulta informatiky a informačných technológií,
Slovenská technická univerzita v Bratislave

Institute of Informatics, Information Systems and Software Engineering,
Faculty of Informatics and Information Technologies,
Slovak University of Technology in Bratislava.

Autor/Submitter: Ing. Igor Stupavský
Institute of Informatics, Information Systems and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava

Školiteľ/Supervisor: prof. Valentino Vranić, PhD.
Institute of Informatics, Information Systems and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava

Oponenti/Reviewers: Ing. Ivana Budinská, PhD.
Ústav informatiky
Slovenská akadémia vied

doc. Ing. William Steingartner PhD.
Katedra počítačov a informatiky
Fakulta elektrotechniky a informatiky
Technická univerzita v Košiciach

Autoreferát bol rozoslaný/Dissertation Thesis Abstract was sent: . . . . . .
(dátum rozoslania/Date of sending)

Obhajoba dizertačnej práce sa bude konať dňa/Dissertation Thesis Defence will be held on ..... o/at ..... h (am/pm) na/at the Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies, Slovak University of Technology in Bratislava (Ilkovičova 2, Bratislava).

.....
prof. Ing. Ivan Kotuliak, PhD.
Dekan/Dean FIIT STU Bratislava

Anotácia

Vo svete, keď sa vo väčšine životných situácií môžeme vo väčšej alebo menšej miere obrátiť na riešenia vo virtuálnom priestore, sa často stáva, že sa nakoniec stanú jediným zdrojom informácií. Pre možnosť sa kvalifikovane rozhodnúť podľa dostupných informácií je nutné môcť sa na takto získané informácie spoľahnúť. Nie je však zaručená integrita týchto informácií ani ich aktuálnosť, ani neutralita obsahu. Na tento spoločenský fenomén vyskytujúci sa vo virtuálnom svete preto treba reagovať na mieste jeho vzniku, nie až potom, ako sa masovo rozšíri. Detekcia antisociálneho správania je predmetom záujmu informatikov. S psychológmi výskumníci častejšie hľadajú príčinu daného správania.

V ideálnom prípade má používateľ dostupné všetky relevantné informácie potrebné na vykonanie kvalifikovaného rozhodnutia. V skutočnom online svete je však nemožné tento stav dosiahnuť pre jednostranné zameranie média, algoritmus filtrujúci obsah podľa skutočných alebo umelo vytvorených preferencií a ďalších faktorov. Zmenou obsahu, prípadne upozornením na podozrivú časť dokážeme do značnej miery znížiť negatívny dosah správy na čitateľa. Môžeme sa snažiť odstrániť čo najviac škodlivých vplyvov, ktoré by zmenili správanie jednotlivca v porovnaní s očakávaným správaním spoločnosti a jednotlivca. Pochopením mechanizmu tvorby antisociálneho správania z oblasti psychológie významne prispeje k obmedzeniu jeho prejavov v prostredí internetu. Z nájdeného riešenia budú mať prospech používatelia, majitelia, ako aj štátne organizácie. Práca poskytuje náhľad do vedecky overených riešení, na základe ktorých je stanovená predikcia predpokladaných tém, ku ktorým budú vznikať jednotlivé formy antisociálneho správania.

Annotation

In a world where, to a greater or lesser extent, we can turn to solutions in the virtual space for most life situations, it often happens that they eventually become the only source of information. To be able to make a qualified decision based on the available information, it is necessary to be able to rely on the information obtained in this way. However, the integrity of this information, its currentity, or the neutrality of the content is not guaranteed. This social phenomenon occurring in the virtual world must therefore be reacted to at the point of its origin, not until it has spread en masse. The detection of antisocial behavior is a subject of interest for informatics researchers. In psychology, researchers are more often looking for the cause of a given behavior.

Ideally, the user has all the relevant information available to make a qualified decision. However, in the real online world, this state is impossible to achieve due to one-sided focus. Media is an algorithm that filters content based on real or artificial preferences and other factors. By changing the content, or warning about a suspicious part, we can reduce the negative impact of the message on the reader to a large extent. We can try to remove as many harmful influences as possible that would change an individual's behavior compared to the expected behavior of society and the individual. Understanding the mechanism of antisocial behavior from the field of psychology will significantly contribute to creating its manifestations limiting the Internet environment. Users, owners, and state organizations will benefit from the solution found. The thesis provides insight into scientifically proven solutions based on which the prediction of the assumed topics is established, which will lead to the emergence of individual forms of antisocial behavior.

Obsah

1 Úvod

V análoch psychologického bádania je osobnosťou, ktorej vplyv rezonuje dodnes, americký psychológ a priekopnícky zakladateľ behaviorizmu (John Broadus Watson). S neochvejným záväzkom vedeckej prísnosti a empirického pozorovania tento vizionársky mysliteľ pretvoril obrysy psychologického myslenia. Tvrdením, že štúdium správania by malo byť základným kameňom psychologického výskumu, boli zatienené princípy introspekcie, čím sa pripravila pôda pre novú éru v psychológii a IT technológiách [1].

Zakladateľ behaviorizmu, ktorý prijal myšlienku, že pozorovateľné akcie slúžia ako základné okná do labyrintových zložitostí ľudskej mysle, ohlasoval zmenu paradigmy. Tento seizmický odklon osvetlil empirické základy ľudského správania a otvoril éru starostlivého experimentovania a kvantifikovateľnej analýzy. Dedičstvo tohto svetla, ktoré je neoddeliteľne spojené s princípmi stimulu a odozvy, sa odráža naprieč oblasťami psychológie a mimo nej, plodí rad teórií a metodológií, ktoré naďalej osvetľujú zložitú tapisériu ľudského správania.

Falošné správy sú vážnou prekážkou v dnešnom informačnom ekosystéme, vrhajú tieň na spoľahlivosť informácií a narúšajú prvotný základ vzdelanostného rozhodovania. Šírenie falšovaného alebo zavádzajúceho obsahu urýchlilo rozšírenie digitálnych platforiem a sociálnych médií, čím sa stierajú rozdiely medzi realitou a fikciou. Manipuláciou s vnímaním verejnosti tento jav spôsobuje spory, narúša dôveru v tradičné médiá a podkopáva demokratickú štruktúru [2].

Cieľom našej práce bolo čo najpodrobnejšie sa oboznámiť s rôznymi spôsobmi a formami antisociálneho správania v online komunitách. Pri štúdiu správania v online komunitách sme dospeli k poznaniu, že táto téma má základ nielen v informatických vedách, ale ide o multidisciplinárny problém. Má tak oveľa širší základ a na jeho správne pochopenie sa treba venovať veľkému množstvu rôznorodých zdrojov. Práve z tohto dôvodu sme sa venovali pochopeniu antisociálneho správania vo veobecnosti.

Antisociálne správanie je prítomné v našej spoločnosti už celú históriu. V súčasnosti, keď väčšina ľudí komunikuje online, je šírenie takéhoto správania oveľa nebezpečnejšie. Môže zasiahnuť a ovplyvniť širokú komunitu používateľov za pár minút či hodín prakticky na celom svete. Internet je tak silný nástroj vhodný na komunikáciu a získavanie relevantných informácií a kontaktu s blízkymi či priateľmi.

1.1 Význam štúdie

Identifikácia potenciálnej zmeny typov antisociálneho správania v bežnom politickom roku a v období pred voľbami, keď ovplyvnenie voliča v správnom čase má potenciál zmeniť výsledok volebného procesu, a teda aj riadenie hospodárstva a presadzovaných zákonov. Správne načasovanie potenciálnej ochrany v správnom čase je tak nevyhnutné na minimalizáciu nežiaducich zásahov.

Návod, ako dopredne získavať potrebné informácie o potenciálnych témach vznikajúcich foriem antisociálneho charakteru, príprava na zníženie hospodárskych škôd na úrovni organizačných jednotiek štátnej správy, až po vyšetrenie vzniknutých incidentov s nasledujúcim vyvodením zodpovednosti aktérov.

Zlepšenie odhaľovania potenciálnych trestných činov veľkého rozmeru veľkej závažnosti s negatívnym dosahom na zdravie a život obyvateľov. Trestné činy veľkého rozsahu sú celospoločensky nebezpečné a vo veľkej miere zahŕňajú terorizmus páchaný buď organizovanými skupinami, alebo jednotlivcami vo forme útokov tzv. “osamelých vlkov”.

1.2 Prístupy navrhnuté v dizertačnej práci

Táto dizertácia prináša metódy a prístupy k modelovaniu, detekcii a predikcii antisociálneho správania vo virtuálnom priestore nezávislú od sociálnych platforiem. Jednotlivé príspevky a nové metódy prezentované v tejto práci môžeme nájsť s ich podrobnejším popisom v nasledujúcich miestach práce:

1. Nový prístup s názvom **Predikcia tém antisociálneho správania** a skratkou **ABtheP** (**A**ntisocial **b**ehavior **the**me **p**rediction) je prezentovaný v sekcii 4.1.
2. Nový prístup s názvom **Predikcia teroristických útokov osamelých vlkov** a skratkou **PredictLWTerr** (**Predict**ing **L**one **W**olf **Terr**orist Outbreaks) je prezentovaný v sekcii 4.2.
3. Nový prístup s názvom **Právny manažment prípadov antisociálneho správania** a skratkou **ABAct** (**A**ntisocial **B**ehavior **Act**) je prezentovaný v sekcii 4.3.
4. Nový prístup s názvom **Zníženie potrebného objemu detegovaného textu pomocou umelej inteligencie a veľkých jazykových modelov** a skratkou **Reduct_text_AI_LLM** (**Reduc**(t)ing the required volume of detected **text** using **A**rtificial **I**ntelligence and **L**arge **L**anguage **M**odels) je prezentovaný v sekcii 4.4.

2 Teória

Antisociálne správanie definujú rôzni autori odlišne. Napríklad L. Selecká a kol. používajú definíciu *"antisociálne správanie (antisocial behaviour) v podobe relatívne širokého spektra veku neprimeraných činov a postojov, ktoré narúšajú očakávania rodiny, sociálne normy, osobné a majetkové práva ostatných, bude pretrvávať aj v neskorších vývinových obdobiach. Zároveň ide o jeden z typických znakov obdobia adolescencie, pre ktoré je charakteristický nárast výskytu prejavov externalizovanej i internalizovanej psychopatológie."* [3], zatiaľ čo napríklad S. Kumar a kol. *"ako šírenie dezinformácií a reakcie antisociálnych používateľov"* [4].

2.1 Formy antisociálneho správania a dôsledky ich vplyvu na jednotlivca a spoločnosť

Nejednotnosť v týchto definíciách antisociálneho správania podtrhuje aj fakt, že z hľadiska miesta pôsobenia tohto protispoločenského správania môžeme uvažovať o antisociálnom správaní v reálnom (fyzickom) svete a v online svete. V rámci reálneho sveta sa vyskytuje v prevažnej miere vo forme hrubej sily a fyzického násilia. V online svete je tento typ správania zastúpený rôznymi aspektmi. Medzi najviac publikované prejavy sem patrí podľa P. Návrata a kol. [5] patria:

1. Falošné a skreslené správy (*angl. fake and biased news*)
2. Falošné recenzie (*angl. fake reviews*)
3. Hoax (*angl. hoax*)
4. Trolovanie (*angl. trolling*)
5. Manipulovanie diskusií (*angl. sockpuppeting*)
6. Zneužívanie komunity (*angl. community abuse*)
7. Vandalizmus (*angl. vandalism*)
8. Záškodnícke správanie hráčov online hier (*angl. griefing*)
9. Kyberšikana (*angl. cyberbullying*)
10. Hejtovanie (*angl. hating*)
11. Spamovanie (*angl. spamming*)

V oblasti, ktorej sa venujeme, nás prioritne zaujímali prejavy antisociálneho správania, ktoré môžeme pozorovať v politickej oblasti. Medzi ďalšie prejavy, ktoré nie sú zahrnuté v predchádzajúcom delení, patrí napríklad:

Propaganda, ktorá bola bežným nástrojom politických strán na ovplyvňovanie verejnej mienky už v časoch pred internetom. V ére informačno-komunikačných technológií však propaganda dostala

možnosť šíriť sa intenzívnejšie a rýchlejšie. V článku [6] sa možno dočítať, že "propaganda ako účinný prostriedok vplyvu sa spája predovšetkým s komunikačnými aktivitami totalitných autoritárskych režimov, ako aj s extrémnymi sociálnymi situáciami". Podkategórie ako čierna a biela nesú označenie podľa cieľa, ktorý chce autor (prípadne šíriteľ) dosiahnuť šírením takejto informácie.

Konšpiračné teórie definuje [7] ako vieru, že rôzne javy sú tajne plánované veľkými organizáciami s cieľom poškodiť, kontrolovať, zavádzať a ovládať spoločnosť. Tieto teórie sú ťažko vyvrátiteľné a ak sú aj dôkazy proti takejto teórii, jej obhajcov to utvrdí v ich viere ešte viac. Preto sme sa týmto teóriám v našej práci nevenovali.

Dezinformácie do tejto práce patria, ale ako vidno na nasledujúcom obrázku, úzko súvisia s fake news. Ako sa však uvádza v [8], samostatný článok, ktorý by dával do súvislosti dezinformácie a fake news chýba.

Satira a paródia má v správach, pomocou ktorých je šírená, podobné znaky ako hoax alebo fake news. Preto je ich odlíšenie do značnej miery ťažké. Tento fakt využívajú tvorcovia antisociálneho obsahu a to tým spôsobom, že svoju prácu označia na konci za satiru. V tomto prípade ide o otvorený problém.

2.2 Tvorba bezpečnostnej politiky z pohľadu rôznych foriem antisociálneho správania

Pri každom spracovaní dát v informatike je potrebné myslieť na bezpečnosť. Vzhľadom na tému našej práce môžeme identifikovať bezpečnosť organizácie pri náchylnosti na nežiaduce reakcie zamestnancov [9, 10] (od harašmentu cez nečestné správanie zamestnancov, až po priemyselnú špionáž) a pomocou vonkajších zdrojov [11], ale aj bezpečnosť z pohľadu zberu informácií zo sociálnych sietí [12] a zachovanie bezpečnej integrity už zozbieraných dát [13], aby mohli byť použité pri trestnom stíhaní jednotlivých osôb podozrivých z páchania trestnej činnosti.

2.2.1 Metodika OSINT

Open Source Intelligence (OSINT) je metodika združujúca získavanie informácií z takzvaných otvorených zdrojov. Podľa [14] približne 80% až 90% údajov, ktoré získavajú spravodajské služby sa získava z verejne dostupných informácií.

S metodikou OSINT bezprostredne súvisia aj pojmy ako SOCMINT, GEOINT a HUMINT:

SOCMINT definujeme napríklad v [15] ako skratku pre inteligenciu zo sociálnych médií, v ktorých pomocou digitálnej forenznej analýzy vieme spracovať tzv. kybernetické profilovanie. V spomínanej štúdii analyzovali možnosť získania dostatočného množstva dát na kybernetické profilovanie aj z relatívne obmedzeného množstva dostupných zdrojov. Analyzovali sociálnu sieť "Minds", ktorá okrem bežných funkcií obsahuje aj reputačné skóre a využíva sa na zber údajov pre odborné analýzy podporované aj skupinou Anonymous [16].

GEOINT je geopriestorová inteligencia, keď sa rôzne získané informácie zasadzujú nie len do časového, ale aj do priestorového logického sledu. Použitie tejto techniky možno nájsť v rôznych inžinierskych alebo sociálnych odboroch. Podľa [17] sa často využíva vo vojenskej oblasti, napríklad Severoamerická národná geointeligenčná agentúra (NGA) [18].

HUMINT sa prekladá ako ľudská inteligencia. Najčastejšie ide o interakciu používateľov, napríklad v [19] analyzujú schopnosť úradníkov v Japonsku riadiť operácie v zahraničí. V [20] poukázali na možnosť nepriamej kontroly politickými stranami, často cez zdieľané texty s nejakým druhom naratívu na sociálnych sieťach.

2.3 Spracovanie prirodzeného jazyka

Všetky prejavy antisociálneho správania v online priestore používajú ako nástroj svojho šírenia text. Analýzou textu sa zaoberajú metódy, ktoré súhrnne označujeme ako spracovanie textu, alebo spracovanie prirodzeného jazyka (angl. Natural language processing, NLP).

Podľa práce [21] môžeme spracovanie prirodzeného jazyka rozdeliť na rôzne typy analýz od fonologickej, cez syntaktickú, až po takzvanú pragmatickú analýzu.

Medzi NLP všeobecne môžeme zahrnúť všetky spôsoby, akými počítač (všeob. informačno-komunikačná technológia) komunikuje s človekom a naopak. Pre potreby nášho výskumu sme sa zamerali na analýzu textu použitého v jednotlivých sociálnych sieťach. Ďalším dôležitým znakom je jazyk analyzovaného textu, ktorým je pre naše potreby slovenský jazyk. Voľba slovenského jazyka bola prirodzená z dôvodu zamerania na politickú scénu v Slovenskej republike.

2.3.1 Tokenizácia

Cieľom tokenizácie je identifikácia jednotlivých slov a vetných hraníc, ktoré môžu slúžiť ako vstup do ďalšieho spracovania. V tomto kroku sme sa tiež snažili zjednotiť spôsob zápisu číslic, diakritiky, interpunkcie, akronymov, symbolov, skratiek a iných významových jednotiek. Táto príprava v sebe zahŕňa kroky ako prepis číslic alebo symbolov. Tento krok je dôležitý v prípade textov stiahnutých z internetu (napríklad FB, Twitter), ktoré často obsahujú množstvo emotikonov a podobne, ktoré by bránili správnej identifikácii a mohli by spôsobiť napríklad nesprávnu tokenizáciu. Autori uvádzajú príklad z analýzy slovenského jazyka: „V slovenských textoch môže byť bodka súčasťou označenia číselného poradia, skratky alebo e-mailovej alebo webovej adresy.“ [22].

Pri príprave na tokenizáciu sa stretávame s pojmom čistý text. Autor profesor Paralič a kol. [23] ho definujú ako „... sekvenciu alfanumerických, interpunkčných, oddeľovacích grafických znakov a niektorých špeciálnych symbolov (napr. %, &, * a pod.). Alfanumerické znaky majú fonetickú a lexikálnu hodnotu, sú teda priamymi nositeľmi obsahu textu. Oddeľovacie a interpunkčné znaky (napr. medzera, tabulátor, bodka, čiarka, pomlčka, zátvorky, lomka atď.) sa používajú na členenie textu.“

2.3.2 Lematizácia a stemovanie

Niektorí autori zaraďujú stemovanie pod špeciálny druh lematizácie [23]. Všeobecne ide o postup odstraňujúci predpony a prípony slov. Z tokenu (slova) získame základ, ktorý nazývame term.

K redukcii jednotlivých slov dochádza podľa [24] rôznymi metódami:

1. metódou slovníka kmeňov a koreňov – výhodou je minimálna chybovosť, nevýhodou rozsiahlosť potrebných slovníkov a špecifickosť odboru
2. odstránením afixov – odstránenie predpôn (prefixov) a prípon (sufixov), potrebný je zoznam prefixov a afixov
3. štatistickými metódami – na základe variet po sebe idúcich písmen v slove, výhodou je nezávislosť od jazyka textu.

Pre slovenský jazyk môžeme nájsť ukážku napríklad v knihe [25]:

$$\{mesto,\ mesta,\ meste,\} \rightarrow mest + 0 \tag{2.1}$$

2.3.3 Parsovanie

Parsovanie, alebo syntaktická analýza je proces, v ktorom popisujeme gramatickú štruktúru pomocou parsovacieho stromu. Parsovacie stromy predstavujú priame závislosti vo vete [26]. Problematickou

sa však stáva veta slovenskom jazyku, keď sa využívajú aj prvky nárečových slov, slov s preneseným významom, ako keď sa pre nejakého politika zavedie prezývka a podobne. Texty na sociálnych sieťach sú často písané takýmto štýlom a parsovanie tak môže mať nižšiu účinnosť v porovnaní s inými metódami NLP používanými na predpovedanie pravdivostnej hodnoty.

3 Cieľ dizertačnej práce

Zvyšujúce sa množstvo (osobného) obsahu na sociálnych sieťach predstavuje základ postupného prechodu do online sveta. Tvoria sa tak dve prepojené osobnosti v jednom používateľovi. Prvou je osobnosť v reálnom svete, kde je každý zviazaný s pravidlami, ktoré treba dodržiavať. Chodiť do práce či do školy a správať sa v zhode s normami spoločnosti. Druhou osobnosťou sa stáva priemet nášho ja v online svete. Vo svete, ktorý poskytuje čiastočnú anonymitu vo forme skrývania sa za monitor počítača, displej mobilného telefónu, vzbudzujúc na prvý pohľad aj určitú ľahkovážnosť konania či beztrestnosť.

Štruktúra obsahu na sociálnych sieťach sa postupne vyvíjala. V začiatkoch išlo o komunitné zdieľanie svojich osobných záležitostí. Neskôr sa do tohto obsahu pridali známe osobnosti z rôznych oblastí a médiá poskytujúce spravodajský obsah.

Trestná činnosť vyskytujúca sa v online svete má svoje zákutia a temné ciele. S množstvom potenciálne zraniteľného osobného obsahu sa zvyšuje aj špecifikácia antisociálneho správania zacielená na vybranú skupinu ľudí podľa pohlavia, náboženstva, orientácie či iného ľubovoľne odlišovacieho znaku.

4 Návrhnuté prístupy

V jednotlivých sekciách tejto kapitoly uvádzame navrhované prístupy z dizertačnej práce, kde môžete nájsť ich definovanie, bližší popis, evauláciu a sumarizáciu získaných výsledkov:

1. Nový prístup s názvom **Predikcia tém antisociálneho správania** a skratkou **ABtheP** (**A**ntisocial **b**ehavior **the**me **p**rediction).
2. Nový prístup s názvom **Predikcia teroristických útokov osamelých vlkov** a skratkou **PredictLWTerr** (**Predict**ing **L**one **W**olf **Terr**orist Outbreaks).
3. Nový prístup s názvom **Právny manažment prípadov antisociálneho správania** a skratkou **ABAct** (**A**ntisocial **B**ehavior **Act**).
4. Nový prístup s názvom **Zníženie potrebného objemu detegovaného textu pomocou umelej inteligencie a veľkých jazykových modelov** a skratkou **Reduct_text_AI_LLM** (**Reduc**(t)ing the required volume of detected **text** using **A**rtificial **I**ntelligence and **L**arge **L**anguage **M**odels).

4.1 Predikcia tém antisociálneho správania (ABtheP)

Dôležitou súčasťou, ktorá je v odborných článkoch často prehliadana, je potreba predikcie. Téma je náročná najmä z dvoch dôvodov. Napríklad v prípade, ak daná správa (obsahujúca prvky antisociálneho správania) už existuje a vďaka externému faktoru sa jej v krátkom časovom úseku dostáva vysoká miera pozornosti. Typicky ide o rýchle aktuálne správy reagujúce na útok, výbuch v objekte s vyšším stupňom dôležitosti a podobne. Druhým dôvodom, prečo je potrebné vedieť včasne predikovať narastajúcu nebezpečnosť konkrétneho tvrdenia bez pravdivostnej hodnoty, môže byť vznik novej falošnej, alebo zavádzajúcej správy, resp. konšpiračnej teórie.

4.1.1 Získavanie informácií zo sociálnej siete

Spoločnosti, ktoré prevádzkujú bezplatné sociálne siete, sú životne závislé od pravidelnej návštevnosti používateľov konzumujúcich spolu so žiadaným obsahom aj reklamný obsah, za ktorý inzerujúce firmy, ale aj organizácie zaplatia. Prevádzkovanie takýchto sietí súkromnými firmami je činnosť s cieľom dosiahnuť zisk v čo najväčšej miere pri zachovaní všetkých legislatívnych opatrení. Tieto organizácie finančne podporujú reklamy aj z dôvodu ich dobrého zacielenia podľa informácií, ktoré vo svojich profiloch uvedú sami používatelia. Keďže v prípade sociálnych sietí ide o maximalizovanie zisku, sú ich používané algoritmy obchodným tajomstvom, a preto nie sú ochotné tieto algoritmy zdieľať ani so svojou konkurenciou, ani s výskumníkmi. to Dôvodom sú aj niektoré aféry spojené so získavaním údajov, ktoré sa stali v minulosti. Príkladom môže byť aj aféra Cambridge Analytica [27, 28] pri ktorej došlo k zneužitiu citlivých údajov používateľov na získanie politickej výhody jedného z kandidátov vo volebnom procese.

4.1.2 Spracovanie údajov s využitím NLP

Spracovanie údajov je v našom návrhu metódy doménovo nezávislé a možno v ňom použiť rôzne metódy, ktoré však musia obsahovať extrakciu hlavnej témy textovej správy. Takúto extrakciu pomocou

umelej inteligencie sme aplikovali napríklad na získané dáta v našom článku [29]. V danom článku sme používali algoritmus, ktorým na účely nami navrhovanej metódy **ABtheP** získame databázu tém vyskytujúcich sa vo falošných správach vo vstupných textoch. Návrh Use case modelu možno vidieť na obrázku 4.1, kde sme ukázali implementáciu našej metódy ABtheP do navrhovanej aplikácie. Výhodou takto vytvorenej aplikácie je možnosť postupného pridávania nových metód s rôznymi nastaveniami parametrov, ktoré tak zabezpečia väčšiu variabilitu detegovaných typov antisociálneho správania.

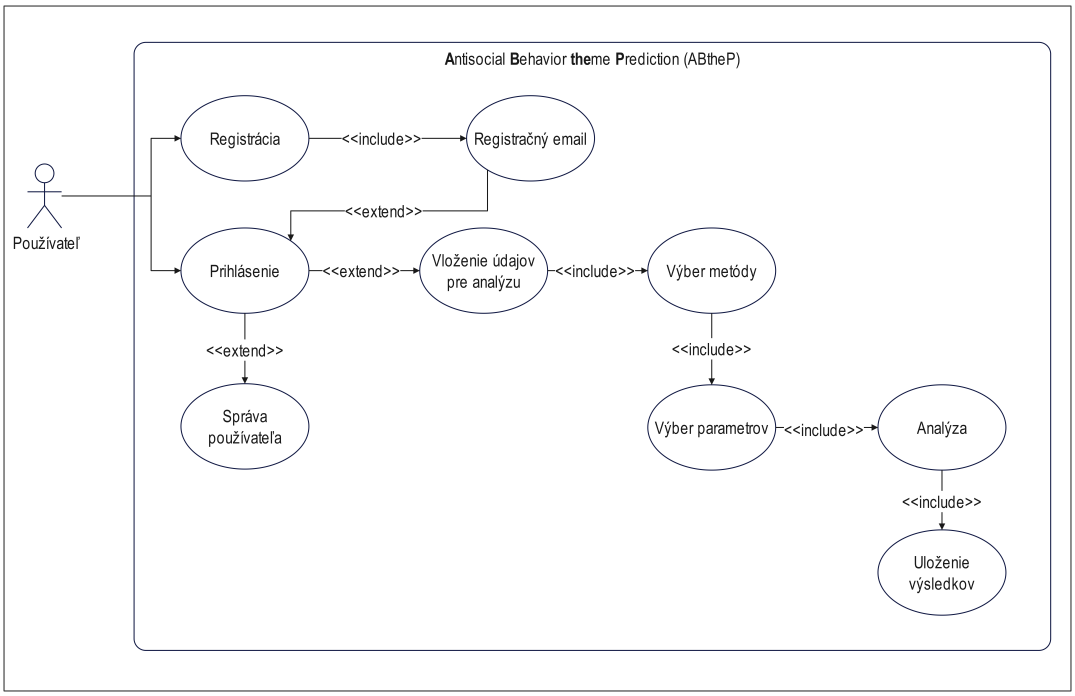


Figure 4.1: Diagram pripadov pouzitia návrhu aplikacie Antisocial Behavior theme Prediction (ABtheP). Zdroj: autorský príspevok

Poslednú čas algoritmu od zistenia tém vo falošných správach (na obrázku návrhu architektúry 4.2 označenej ako "Extrakcia kľúčových slov") sme využili na overenie, či boli vstupné analyzované texty správne identifikované ako falošné správy pomocou metódy citovo zafarbených slov, ktorým sa venujeme v sekcii 4.2.1 našej práce.

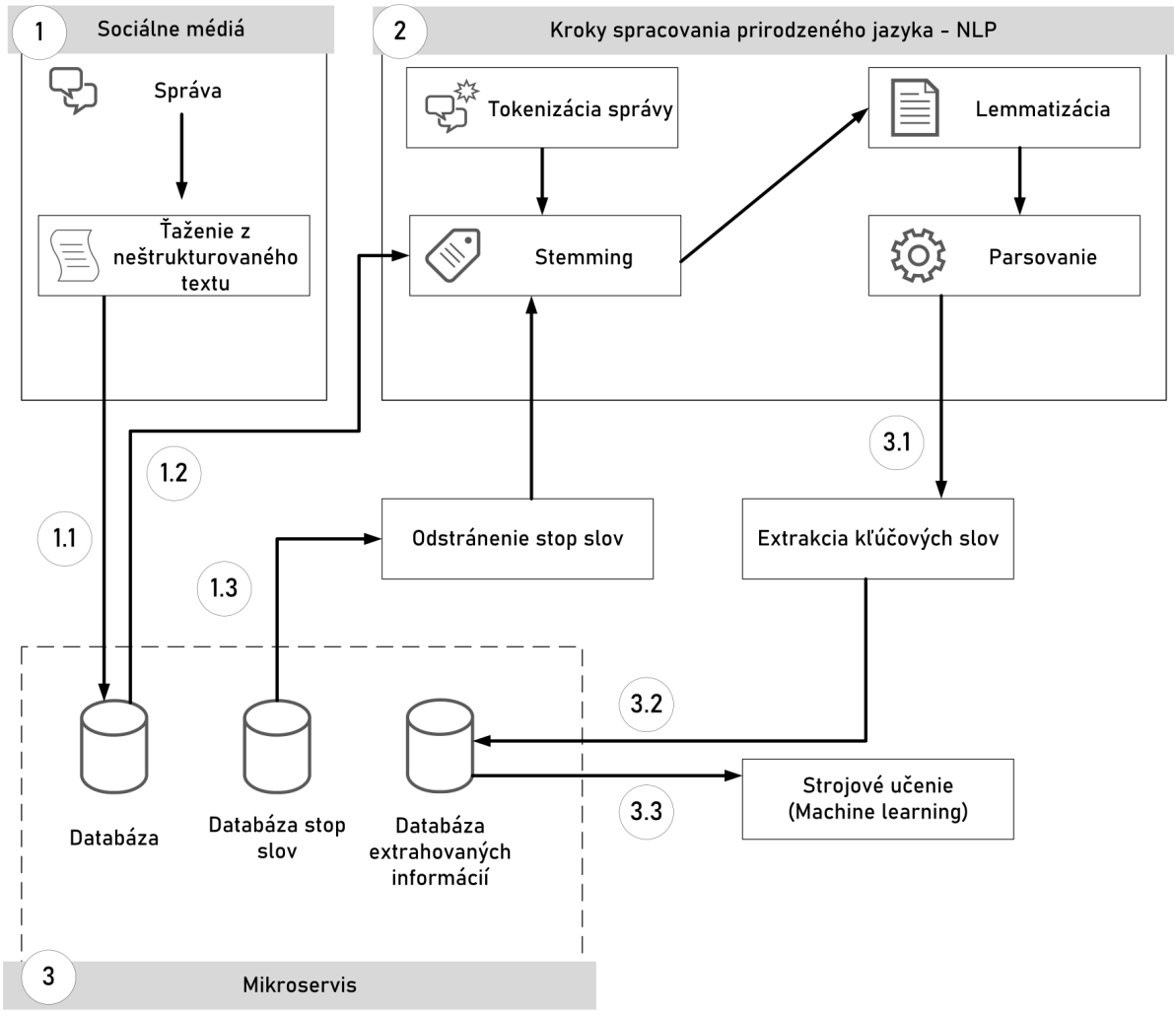


Figure 4.2: Návrh diagramu architektúry. Zdroj: autorský príspevok publikovaný v [2].

4.1.3 Evaluácia

Na overenie našej metódy sme využili formu testovania na vzorovom anotovanom datasete FakeNewsNet [30] pôvodne autormi článku [31] získaného dolovaním údajov z Twitter účtu stránky PolitiFact [32] a z už neexistujúcej stránky GossipCop [33]. K dispozícii sme mali v datasete binárne klasifikované súbory opísané v tabuľke 4.1.

Table 4.1: Početné zastúpenie záznamov v datasete FakeNewsNet. Zdroj: autorský príspevok

Súbor	Početnosť v [ks]	Početnosť v [%]
politifact_fake.csv	432	1.86
politifact_real.csv	624	2.69
gossipcop_fake.csv	5323	22.95
gossipcop_real.csv	16817	72.50

Pri kontrole okometrickou metódou sme si všimli, že vstupné textové dáta obsahujú rôzne jazyky (angličtina, čínština, španielčina), ktorými hovorili jednotliví používatelia zanechávajúci svoje správy. Potrebné správy sme automaticky preložili pomocou nástroja $GoogleTranslator()$ z Python z rovnomen-

ného balíka $GoogleTranslator$ do angličtiny. Odstránili sme stop slová, ktoré nám nepridávali žiadnu dodatočnú významovú hodnotu a aplikovali $nltk.porter.PorterStemmer()$ redukciu na slovný kmeň v NLP známy ako stemming. Dôležitou súčasťou spracovania bolo použitie umelej inteligencie vo forme BERT modelu trénovaného na dátach z CNN Daily Mail. Takto získaný zoznam informácií o ktorých sa v sledovanom období písalo na daných stránkach, vieme uložiť spolu s ich početnosťou do databázy a priradiť im časovú stopu spustenia analýzy.

Opakovaným spúšťaním na reálnych dátach získame pravidelne aktualizovanú znalostnú databázu údajov v štruktúre, ako je prezentovaná v tabuľke 4.2 vo forme základných atribútov. Atribút Id obsahuje jedinečný identifikátor vo forme UUID verzie 4, nasleduje samotná téma tvoriaca znalosť v atribúte $Tema$ s počtom výskytov danej témy v atribúte $Pocetnost$ a $Datum$ typu DateTime, aby bola zachovaná časová informácia. Ďalšou skupinou atribútov sú riadiace atribúty, ktoré jednoznačne identifikujú používateľa (atribút $PouzivatelId$) vo forme UUID na identifikáciu vlastníka údajov nasledované atribútmi na štandardné CRUD operácie (angl. Create Read Update Delete). V prípade atribútu $DeletedAt$ navrhujeme vložiť časový údaj v ďalekej budúcnosti (napr. dátum a čas v roku 2999). Ak by používateľ chcel záznam vymazať, tento atribút sa označí príslušným dátumom a časom zmazania. Samotný záznam však v databáze ostane zachovaný, iba pre používateľa sa stane neviditeľným. Tento postup zachovávame pre potreby spätnej kompatibility zobrazených údajov. Pre používateľa v role administrátora však bude záznam aj naďalej viditeľný aj s údajom o dátume a čase jeho vymazania používateľom.

Table 4.2: Základná štruktúra navrhovanej znalostnej databázy. Zdroj: autorský príspevok

Atribút	Typ
Id	UUID
Tema	Text
Pocetnost	Int
Datum	DateTime
PouzivatelId	UUID
CreatedAt	DateTime
UpdatedAt	DateTime
DeletedAt	DateTime

4.1.4 Sumarizácia

Evaluáciou sme dokázali úspešnosť nášej navrhovanej metódy na vzorovom datasete, ktorý vstupoval do celého procesu vo forme dávkového súboru a tak emuloval vstupné dáta v určitom časovom okamihu tvoriace jednu skupinu tém antisociálneho správania. Opakovanou analýzou dát získame časový rad údajov pre konkrétnu tému, z ktorej sme schopní metódami predikcie dokázať predpovedať nárast s určitou pravdepodobnosťou meniacou sa s použitou prediktórnou metódou. Celú architektúru navrhnutého riešenia sme riešili ako architektúru založenú na mikroservisoch pre potreby škálovateľnosti jednotlivých zúčastnených zložiek. Tie tak tvoria samostatnú jednotku a v prípade poruchy alebo nedostatočného výkonu sa dá daná súčasť riešenia upraviť podľa aktuálnych potrieb.

Naše riešenie sme postavili na prevažnej miere na nami publikovaných riešeniach nachádzajúcich sa v článku [2], ktorý sa zaoberá samotnou metódou na vzorovom datasete a v článku [34], v ktorom sme sa venovali získaniu vstupných dát do navrhnutej metódy z platformy sociálnej siete Facebook.

4.2 Predikcia teroristických útokov osamelých vlkov (PredictLWTerr)

Prezentované typy protispoločenského správania majú v reálnom svete svoje následky a možné postihy podľa definovaných pravidiel slušného správania alebo podľa zákonmi, či medzinárodnými zmluvami stanovených pravidiel. Po prekročení takýchto pravidiel nastupujú zákonné možnosti trestu vzhľadom na definovanie ich závažnosti. V zákone číslo 47/2024 Z. z., ktorým sa dopĺňa zákon č. 300/2005 Z. z. trestný zákon v znení neskorších predpisov sú v § 140b Trestné činy terorizmu [35] definované tieto závažné činy proti zákonu. Nami analyzovaný trestný čin sa všeobecne nazýva útokom "osamelých vlkov" pre spôsob jeho vykonania, kde sa zdanlivo neškodný jedinec dopustí samotného aktu terorizmu, alebo jeho prípravy v rámci definovaného zákona.

4.2.1 Citovo zafarbené slová

V rôznych článkoch bola použitá hypotéza, že správy, ktoré boli vytvorené ako klamlivé, využívajú vo väčšej miere citovo zafarbené (expresívne) výrazy. Preto, ak podrobíme skúmaniu jednotlivé správy na tento druh slov, môžeme s určitou mierou predpokladať ich pravdivosť.

Detekciu citovo zafarbených slov sme tvorili v programovom prostredí jazyka Python, ktoré sa v súčasnosti na tieto formy spracovania textu využíva. Samotné správy boli uložené v JSON formáte.

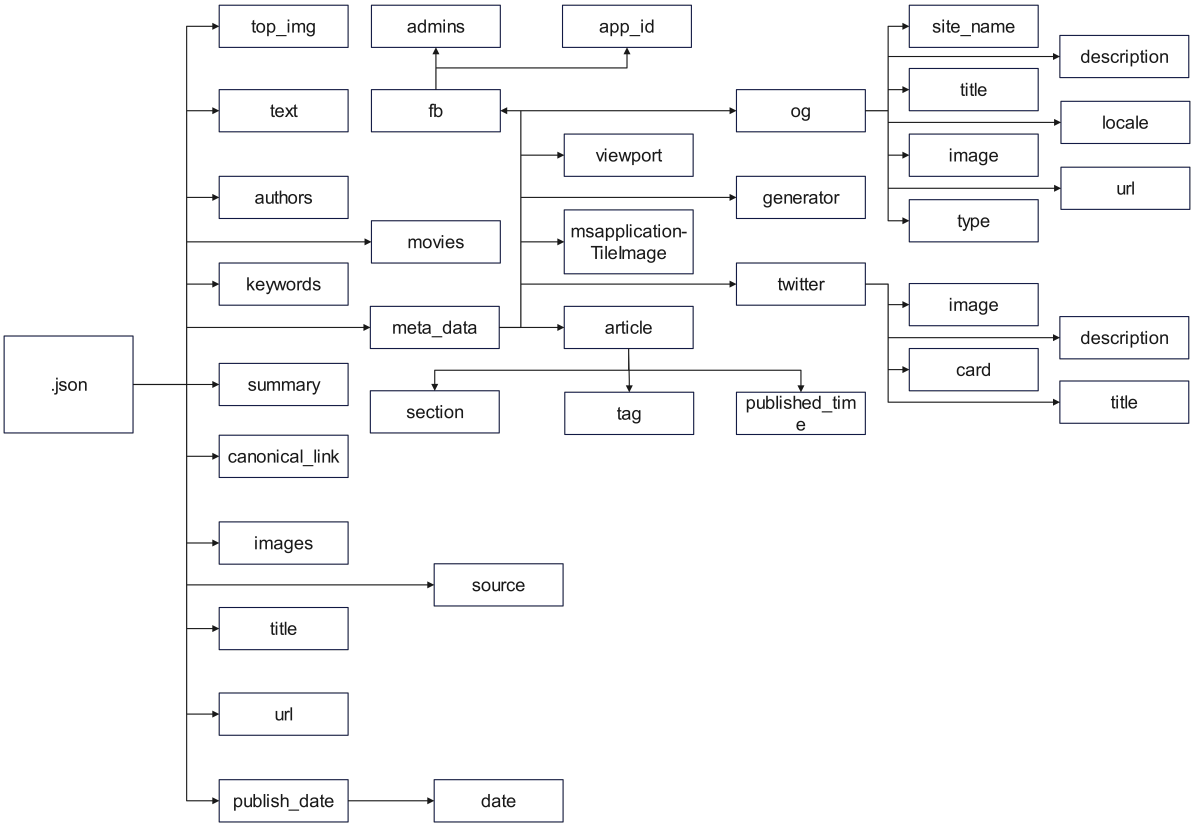


Figure 4.3: Štruktúra analyzovanej správy v JSON formáte. Zdroj: autorský príspevok.

Výsledné hodnoty z tohto experimentu po validácii dosiahli úspešnosť 76.32% pozitívnej identifikácie fake news a 68.00% real news. Týmto experimentom sme dosiahli podobné výsledky ako výskumníci z Univerzity v Santa Barbare [36] vo svojej práci s iným datasetom 85,5% pre SeekingAlpha a 76,2% pre StockTwits.

Dataset, ktorý sme používal v rámci tohto experimentu, je prístupný na [31, 37–39] a obsahoval údaje zozbierané z obľúbenej sociálnej platformy Twitter [40]. Typická štruktúra záznamu (zobrazená na

uvedenom obrázku) obsahovala:

- “top_img” – ide o typický hlavný obrázok statusu na sieti Twitter
- “text” – obsahuje základný text statusu
- “authors” – identifikuje používateľské meno pridávajúceho, ktoré aktuálne používa na sociálnej sieti
- “keywords” – identifikuje hlavné slová
- “meta_data” – obsahuje informáciu o meta údajoch, ktoré sa zaznamenávajú a závisia od typu príspevku a od platformy, prípadne od rôznych prepojení na iné miesta v rámci siete internet (napr.: prepojenie na inú sociálnu sieť Facebook).
- “canonical_link” – obsahuje odkaz na internetové stránky, ktoré sa v statuse spomínajú
- “images” – obrázky, ktoré sú publikované v rámci odkazu z príspevku
- “title” – názov príspevku
- “url” – hlavná URL adresa príspevku
- “summary” – zhrnutie, ale vo väčšine príspevkov v rámci datasetu neprítomné
- “movies” – ostatné pripojené súbory (video súbor a pod.)
- “publish_date” – dátum publikovania príspevku (príklad: “$date”: 1452628948000)
- “source” – zdroj

Každý záznam v použitom datasete bol rozdelený na fake news a real news a bolo tak možné použiť cross validáciu na overenie identifikácie záznamu. Pri identifikácii sme používali binárnu klasifikáciu “fake news” a “real news”.

Slovník, ktorý sme používali bol „NRC Emotional Lexicon“ [41]. Obsahoval 14182 záznamov.

Podrobnejšie problematike citovo zafarbených slov sme sa venovali v našom článku, ktorý je možné nájsť v [42].

Návrh na detegovanie osamelých vlkov vychádza z definovania hypotézy, žè navrhneme aplikáciu pre potreby vyšetrovacích zložiek použiteľné pri súdnych pojednávaniach na odhaľovanie potenciálnych teroristických činov páchaných osamelými vlkmi. Jej návrh sme tvorili spolu s naším diplomantom Róbertom Fajdom a riešenie vyústilo do vypracovania diplomovej práce a pripravovaného vedeckého článku [1]. Pri analyzovaní zozbieraných dát sme používali analýzu sentimentu, emócie, hate speechu, nevhodného slovníka a irónie príspevku zo sociálnej siete Twitter.

[1] Táto metóda je založená na časti z pripravovaného článku [43], kde bude môj príspevok 50%.

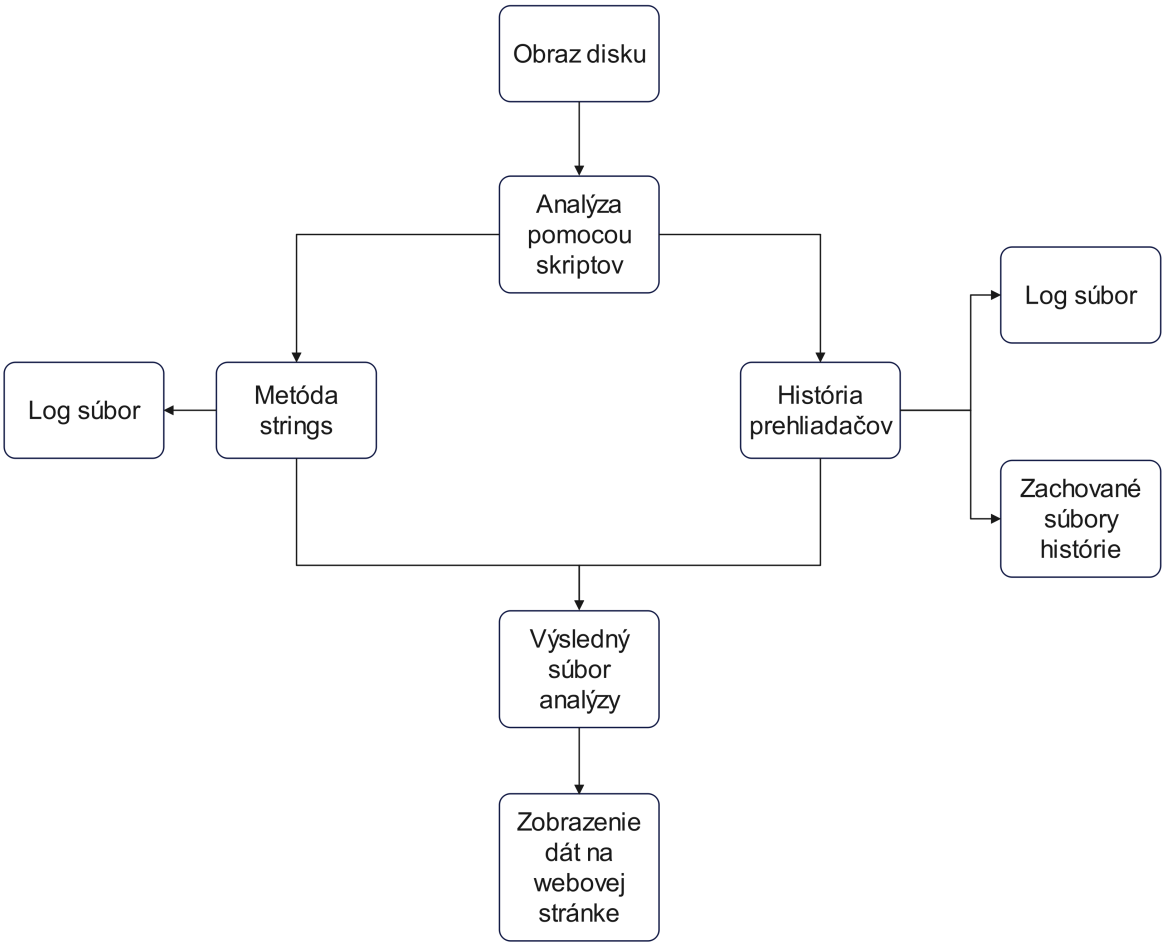


Figure 4.5: Návrh systému forenznej analýzy. Zdroj: [44]

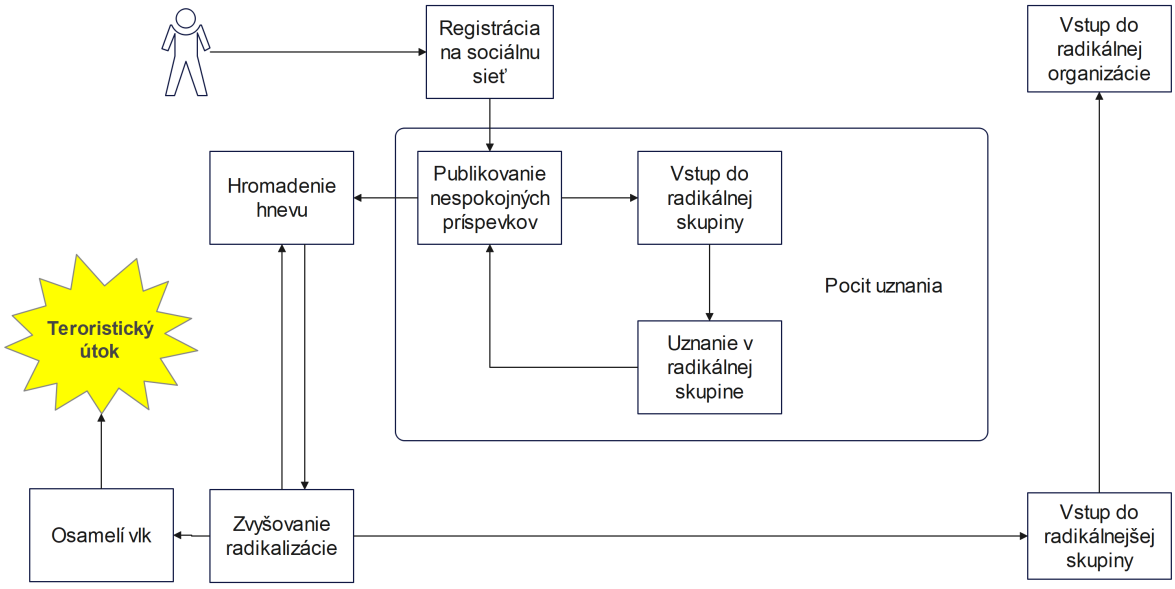


Figure 4.4: Princíp radikalizácie osamelích vlkov na sociálnych sieťach. Zdroj: autorský príspevok

4.2.2 Evaluácia

Na overenie našej metódy sme využili formu testovania na dvoch vzorových Twitter účtoch, z ktorých jeden simuloval analyzovaného používateľa a druhý účet forenzného analytika pre potreby testovania získavania dát cez Twitter API. S testovacím používateľom sa vykonali typické úkony, ktoré odzrkadľovali bežné využívane sociálnej siete. Ako príklad testovaných úkonov si možno predstaviť odoslanie súkromnej správy, zdieľanie obrazovej prílohy či lajknutie statusu inému používateľovi a podobne.

V druhom kroku náš systém po prihlásení do siete po korektnom získaní všetkých potrebných dát a použití vybranej hašhovacej funkcie správne identifikoval použitie statusov, kontaktovanie medzi používateľmi, ako aj zachytenie zasielaných správ.

Tretím krokom bola analýza získaných dát (v tabuľke 4.3), na ktorú sme využili techniky NLP. Pri štvrtom kroku sme vizualizovali získané informácie na ľahšiu interpretáciu výsledkov.

Table 4.3: Údaje získané metodikou OSINT. Zdroj: [43, 45]

Parameter	Hodnota parametru
alt_text	text v prípade nenačítania média
duration_ms	dĺžka
height	šírka
media_key	ID
preview_image_url	link na náhľad
type	typ médií
url	link na médium
variants	varianty médií
width	dĺžka

Následne sme na získaných textoch z Twitter statusov vykonali analýzu a to konkrétne sentimentu, základnej emócie, hate speechu, nevhodného slovníka a irónie. Takto sme vytvorili základný balík analýz, ktoré sa môžu pomocou definovania nových metód naďalej rozširovať. V reálnom použití by sme vybrané metódy nahradili jednotlivými mikroservismi, ako v návrhu našej metódy s názvom **A**ntisocial **b**ehavior **the**me **p**rediction a skratkou **ABtheP** prezentovanej v tejto práci.

Následne sme s naším diplomantom Róbertom Fajdom [45] vykonali vizualizáciu získaných dát prezentovanú v pripravovanom článku [43], kde je náš autorský príspevok 50%. Výsledky analýzy pre vzorového Twitter analyzovaného používateľa možno vidieť v tabuľke 4.4, ktorá ukazuje že sme v našom modelovom prípade získali predpokladané dáta.

Table 4.4: Porovnanie analyzovaných metód na vzorovom Twitter používateľovi. Zdroj: autorský príspevok citovaný v prácach [43, 45]

Analyzovaná metóda	Parameter	Hodnota parametru [%]
sentiment	neutrálna	86
	pozitívna	10
	negatívna	4
základná emócia	hnev	51
	radosť	40
	smútok	8
hate speech	nie hate	100
	hate	0
nevhodný slovník	nie ofenzívna	100
	ofenzívna	0
irónia	nie irodnické	75
	ironické	24

Pri získavaní dát sme využili metodiku OSINT, ktorej sme sa teoreticky venovali sme na zabezpečenie integrity získaných dát použili techniku hašovacích funkcií a preto bol na analýzu vytvorený aj logovací súbor s konkrétnym záznamom prístupov. Anonymizácia dát počas trvania experimentu neprebiehala z dôvodu zmyslu zachovania dát pre použitie v prípade súdneho konania, kde jednotlivé oprávnené štátne orgány činné v trestnom konaní pristupujú k citlivým údajom a môžu s nimi pracovať podľa zákona č. 171/1993 NRSR o Policajnom zbore [46] v aktuálnom znení v štvrtej hlave, kde sa definuje spracúvanie informácií Policajným zborom.

Získané výsledky ukazujú údaje platné v konkrétnom časovom okamihu. Opakovaným automatizovaným spustením analýzy dokážeme získavať údaje o aktuálnom stave sledovaného používateľa so zabezpečením integrity získaných dát. Každé zavolanie príslušnej metódy je označené príslušným štandardizovaným časovým údajom podľa ISO 8601 [47] a zaznačené aj v logovacom súbore. Takýmto posupom získame automatizovane a legislatívne korektne zaznamenaný časový rad údajov, ktoré môžeme využiť na následnú extrapoláciu údajov a získať predikciu dát v určitom časovom okamihu v budúcnosti. Nastavenie jednotlivých upozornení by pri podozrivých osobách boli definované príslušnými oprávnenými orgánmi štátnej správy s príslušnou citlivosťou na sklon strmosti predikcie zmien hodnôt jednotlivých atribútov.

4.2.3 Sumarizácia

Sekcia predstavuje základnú techniku identifikácie citovo zafarbených slov, ktorá sa však počas našej analýzy ukázala ako maximálne efektívna, pretože všetky výskumy podporujú hypotézu, že základnou technikou antisociálneho správania je útočenie na citovú zložku prijímateľa správy. Úspešnosť tejto detekcie sme dokázali aj nami vykonaným experimentom s úspešnosťou 76.32% v akademických podmienkach. Preukázali sme vplyv typu antisociálneho správania (v našom vzorovom prípade falošných správ) naprieč celou spoločnosťou v rámci jednotlivých výrobných odvetví v prostredí národného hospodárstva.

Navrhli sme našu metódu **PredictLWTerr** slúžiacu na predikciu možnosti trestných činov takzvaných "osamelých vlkov" v prostredí a dodržaní štandardov digitálnej forenznej analýzy.

4.3 Právny manažment prípadov antisociálneho správania (ABAct)

Problematika antisociálneho správania je inkorporovaná do všetkých oblastí národného hospodárstva a nevyhli sa mu ani jednotlivé súčasti orgánov štátnej správy. Štát potrebuje ochrániť tak kľúčové zložky svojej správy, ako aj strategické súčasti hospodárstva. Vzhľadom na požiadavky Európskej komisie [48] je žiaduce implementovanie ochrany pred hybridnými útokmi zo strany externých subjektov do ktorých tento typ správania patrí.

Návrh na vytvorenie postupu krokov [2] na minimalizovanie škôd v štátnej správe falošnými správami vychádza z definovania hypotézy dizertačnej práce, že definovaním postupu na predchádzanie hospodárskym škodám pri prejavoch rôznych foriem antisociálneho správania pre štátnu správu a rôzne odvetvia budeme vedieť znížiť ich vplyv na fungovanie štátu.

1. Implementáciou riešenia našej metódy **ABtheP** zistíme kľúčové slová (témy antisociálneho správania).

2. Pomocou predikcie časovej zmeny kľúčových slov (definovaných na konci sekcie 4.2.2) zisťujeme najpravdepodobnejšiu tému novej falošnej správy na definované časové okno (bude definované pre vybrané pomocou možného reakčného času na ochranu daného elektronického systému).

3. Zabezpečíme zvýšenú mieru ochrany kľúčových oblastí dotknutých elektronických systémov.

4. Vytvoríme oznámenie / krátku správu vysvetľujúcu reálne okolnosti predmetnej témy.

5. Zdieľame oznámenie na všetkých médiách dotknutých orgánov štátnej správy.

6. V prípade potreby informujeme v celoštátnych médiách, ako je televízne spravodajstvo a podobne.

7. Pomocou časovej zmeny identifikujeme paralelne metódami OSINT približnú oblasť vzniku falošnej správy.

8. Metódami OSINT opísanými v tejto práci zabezpečíme potrebné forenzné dôkazy.

9. Vypracujeme logickú interpretáciu zistených dôkazov.

10. Informujeme orgány činné v trestnom konaní a prípadne aj verejnosť o riešení.

[2] Táto sekcia je založená na časti z pripravovaného článku [49], kde je môj príspevok 100%.

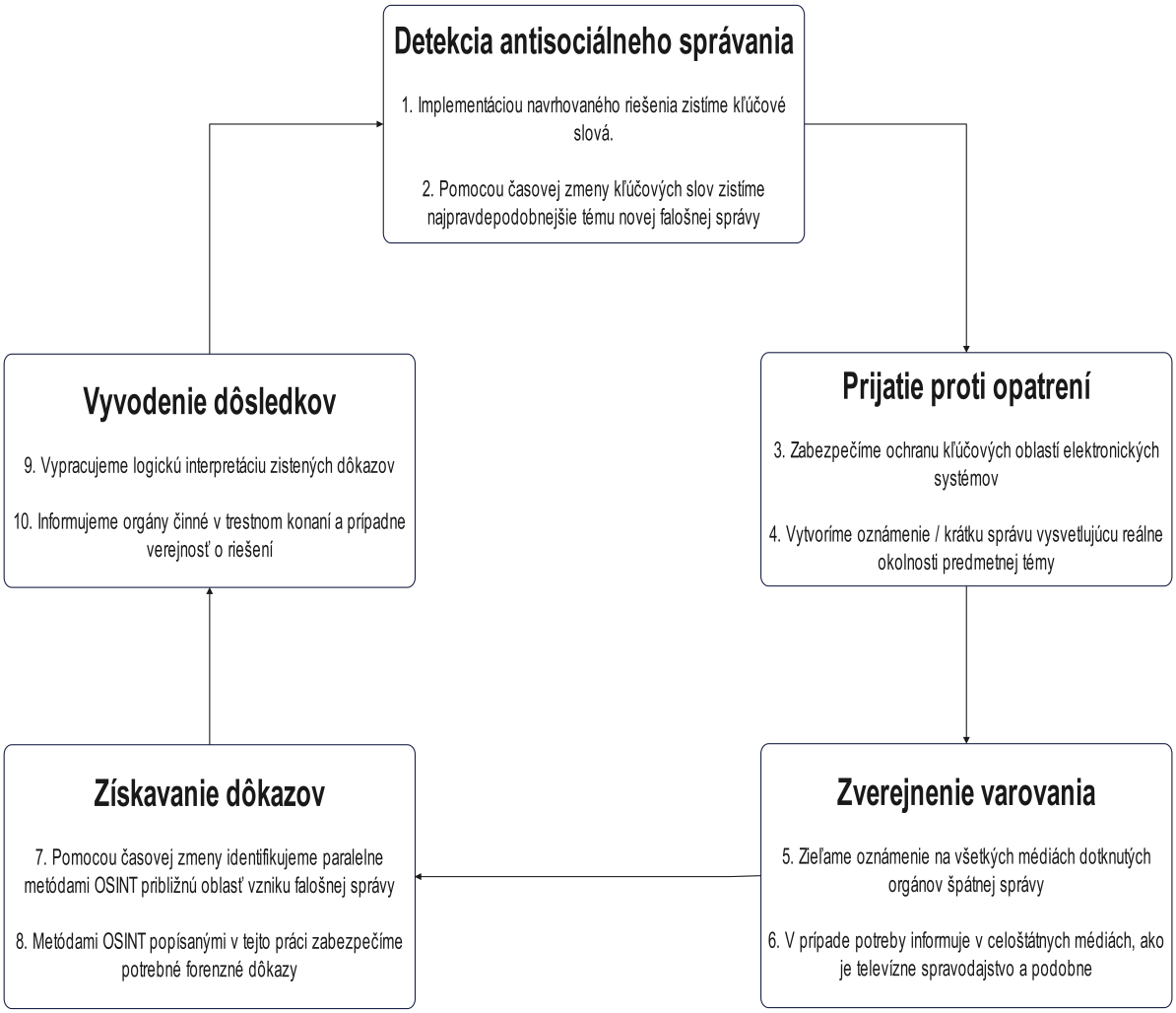


Figure 4.6: Špirála riešenia. Zdroj: autorský príspevok.

4.3.1 Evaluácia

Na evaluáciu funkčnosti sme vytvorili 3 samostatné testovacie scenáre s testovacími krokmi pre základnú funkcionalitu systému. Zoznam testovacích scenárov vyplýva z tabuľky 4.5.

Table 4.5: Testovací scenár. Zdroj: autorský príspevok

Funkčný testovací scenár	**Testovací prípad**
Verifikovať funkcionalitu detekcia témy antisociálneho správania	Detegovať nové témy antisociálneho správania
Verifikovať funkcionalitu odoslanie na analýzu	Systém odošle analyzované údaje do expertízneho ústavu
Verifikovať funkcionalitu odoslanie príslušným štátnym orgánom	Systém odošle analyzované údaje s výsledkom analýzy na príslušné orgány štátnej správy

Testovací prípad s popisom krokov v tabuľke 4.6 popisuje verifikáciu požadovanej funkcionality na

detekciu tém falošných správ.

Table 4.6: Kroky používateľského scenára 1. Zdroj: autorský príspevok

Krok	Detail kroku	Očakávaný výsledok	Aktuálny výsledok
1	Používateľ načítal stránku so systémom	Stránka sa zobrazila	OK
2	Používateľ zadal používateľské meno	Vo formulári v poli prihlasovacie meno sa zobrazí zadaný text	OK
3	Používateľ zadal používateľské heslo	Vo formulári v poli prihlasovacie meno sa zobrazí zástupný znak podľa počtu zadaných písmen	OK
4	Používateľ klikol na tlačidlo “Prihlásiť”	Aplikácia verifikovala používateľa a načítala domovskú obrazovku	OK
5	Používateľ klikol na tlačidlo “Nové hľadanie tém”	Systém spustil detekciu a zobrazil výsledky	OK
6	Používateľ klikol na “Zapísať časový rad”	Systém vytvoril nový záznam do tabuľky temy_v_case	OK
7	Používateľ klikol na “Analyzuj časový rad”	Systém overil rozdiely v rovnakých témach rôznych časových radov a upozornil na výsledok	OK
8	Používateľ klikol na tlačidlo “Nahlásiť”	Systém vytvoril report a zasla ho na zvolenú e-mailovú adresu	OK
9	Používateľ získal odpoveď a klikol na tlačidlo “Vytvor externú správu”	Systém otvorí modálne okno s textovým poľom	OK
10	Používateľ napísal textovú správu a klikol na tlačidlo “Odoslať”	Systém odošle vytvorenú správu na definovaný zoznam e-mailových adries	OK

Testovací prípad s popisom krokov v tabuľke 4.7 popisuje verifikáciu odoslania údajov na analýzu do expertízneho ústavu.

Testovací prípad s popisom krokov v tabuľke 4.6 popisuje verifikáciu odoslania analyzovaných dát s výsledkom z expertízneho ústavu príslušným orgánom štátnej správy na ďalšie riešenie.

Z uvedených testovacích krokov s očakávaným výsledkom možno vykonať testovanie za prítomnosti zástupcov orgánov štátnej správy (ako zákazníkov) bez nutnosti povolenia od spoločností vlastniacich sociálne siete, keďže dáta boli získané metodikou OSINT. Je však celosvetovo vhodné, aby sa prijalo nariadenie na národnej a medzinárodnej úrovni aspoň o odporúčanej implementácii nášho riešenia. S

Table 4.7: Kroky používateľského scenára 2. Zdroj: autorský príspevok

Krok	Detail kroku	Očakávaný výsledok	Aktuálny výsledok
11	Používateľ klikol na tlačidlo "OSINT"	Systém analyzoval výskyt témy v časových radoch a vytvoril strom	OK
12	Používateľ klikol na koreň stromu	Systém otvorí modálne okno a zobrazí údaje o prvotnom zázname	OK
13	Používateľ klikne na tlačidlo "Odoslať na analýzu"	Systém odošle potrebné údaje na analýzu do expertízneho ústavu	OK

Table 4.8: Kroky používateľského scenára 3. Zdroj: autorský príspevok

Krok	Detail kroku	Očakávaný výsledok	Aktuálny výsledok
14	Používateľ dostane odpoveď z expertízneho ústavu a klikne na "Odoslať dôkazy"	Systém odošle údaje s výsledkom z expertízneho ústavu na príslušné orgány štátnej správy	OK
15	Používateľ klikne na tlačidlo "Odhlásiť"	Systém používateľa odhlási a zobrazí prihlasovaciu obrazovku	OK

väčším počtom používateľov získame lepšie dáta, ktoré nám umožnia lepšie zacielenie vzniku na vznik falošnej správy a následne identifikáciu šíriteľa s potrebnými forenznými dôkazmi v prípade súdneho procesu.

4.3.2 Sumarizácia

Definovali sme novú metódu postupu **ABAct**, jej implementáciou bude môcť štátna správa naplniť požiadavky Európskej komisie, spolupracovať na nadnárodnej úrovni s jednotlivými spojencami a zároveň získať včasnú detekciu správ so spornou pravdivostnou hodnotou. Zistením predtým neznámych informácií tak bude môcť štátna správa včasne prijímať opatrenia na zabránenie hospodárskym škodám a ochrániť obyvateľstvo.

4.4 Zníženie potrebného objemu detegovaného textu pomocou umelej inteligencie a veľkých jazykových modelov (educt_text_AI_LLM)

Sekcia sa venuje problematike potreby analýzy veľkého objemu textu získaného z príspevkov publikovaných na sociálnej sieti členmi kontroverznej komunity a tým jeho stále stúpajúcej náročnosti na dostupnú IKT. Aplikovaním navrhovaného postupu dochádza využitím samotných vlastností príspevkov na sociálnych sieťach k značnej redukcii množstva potrebného textu vstupujúceho do samotnej analýzy. Následne s použitím prvkov AI a LLM sme boli schopní vyvinúť metódu s názvom **Reduc**(t)ing the required volume of detected **text** using **A**rtificial **I**ntelligence and **L**arge **L**anguage **M**odels a skratkou

Reduct_text_AI_LLM.

Analyzované správy vo vzorovom datasete, ktorý sme využili počas návrhu a následne aj pri evaluáci mali binárnu anotáciu. Samotná anotácia môže byť ovplyvnená subjektívnym názorom anotátora a takéto chybné dáta by mohli skresľovať výsledky nášho modelu. Z toho dôvodu navrhujeme implementovať ďalšiu vrstvu, ktorá bude automaticky určovať na základe vybranej metódy pravdivostnú hodnotu analyzovanej správy.

[3] Viaceré výskumné práce vyvodili záver, že častým spôsobom šírenia falošných správ je iba zmena nadpisu správy. Následne môže byť obsah správy prakticky identický. Tento trend vychádza z toho predpokladu, že sociálne siete, ktoré sú často jediným zdrojom správ pre mladšiu populáciu zobrazujú pri zdieľaní konkrétnej správy iba profilový obrázok a nadpis príspevku. Je dôležité tento proces správne identifikovať a následne experimentálne overiť.

Pri výbere použitého datasetu sme vychádzali z požiadavky značeného datasetu pre potreby cross validácie a druhou základnou požiadavkou bola oblasť textov v datasete. Oblasťou, ktorej sa venujeme v našej práci, je oblasť politiky. Dataset, ktorý vyhovoval našim požiadavkam sa nakoniec stal dataset použitý napríklad aj v práci [31].

Table 4.9: Počty zastúpenia reálnych a falošných správ v datasete FakeNewsNet. Zdroj: [29]

	Real	**Fake**
PolitiFact	624	432
GossipCop	16817	5323

Dataset sme podrobili EDA analýze pre pochopenie štruktúry dát.Následne sme dataset podrobili procesu, ktorého postup vidieť na v tabuľke 4.10

Table 4.10: Postup procedúr pri redukcii množstva textu na analýzu. Zdroj: [29]

Procedúra	**Reálna správa v datasete**	**Falošná správa v datasete**
Originálny text nadpisu	"Donald Trump exaggerates when he says China has 'total control' over North Korea"	"Barack Obama Tweets SICK Attack On John McCain, Says He Should Have Died"
Odstránenie stop slov	"Donald Trump exaggerates says China 'total control' North Korea"	"Barack Obama Tweets SICK Attack On John McCain, Says He Should Have Died"
Stemming	"donald trump exagger say china 'total control' north korea"	"barack obama tweet sick attack on john mccain, say he should have die"
Sumarizácia pomocou AI	"says china has 'total control' of North Korea"	"Barack Obama tweets sick attack on john mccain, says"
Hodnotenie podľa emócií	"neg": 0.0, "neu": 1.0, "pos": 0.0, "compound": 0.0	"neg": 0.507, "neu": 0.493, "pos": 0.0, "compound": -0.8834
Výsledná emócia	Neutrálne tvrdenie	Negatívne tvrdenie

[3]Táto sekcia je založená na časti z publikovaného článku [29], kde je môj príspevok 50%.

4.4.1 Evaluácia

Následne sme vybrali hodnotenie podľa základných emócií (pozitívne, negatívne alebo neutrálne) a zakreslili sme polohy jednotlivých správ do normalizovaného grafu. Takéto rozloženie základných emócií vidíme pre reálne správy na obrázku 4.7. Pre falošné správy na obrázku 4.8.

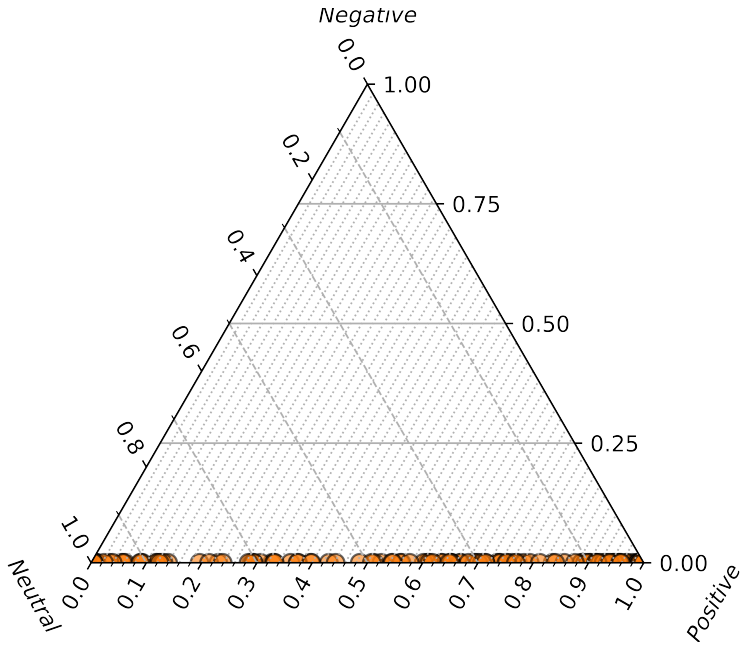


Figure 4.7: Rozloženie základných emócií v nadpisoch časti reálne správy z datasetu FakeNewsNet. Zdroj: autorský príspevok publikovaný v [29]

Reálne správy v datasete FakeNewsNet v politickej oblasti podľa prevládajúcej emócie

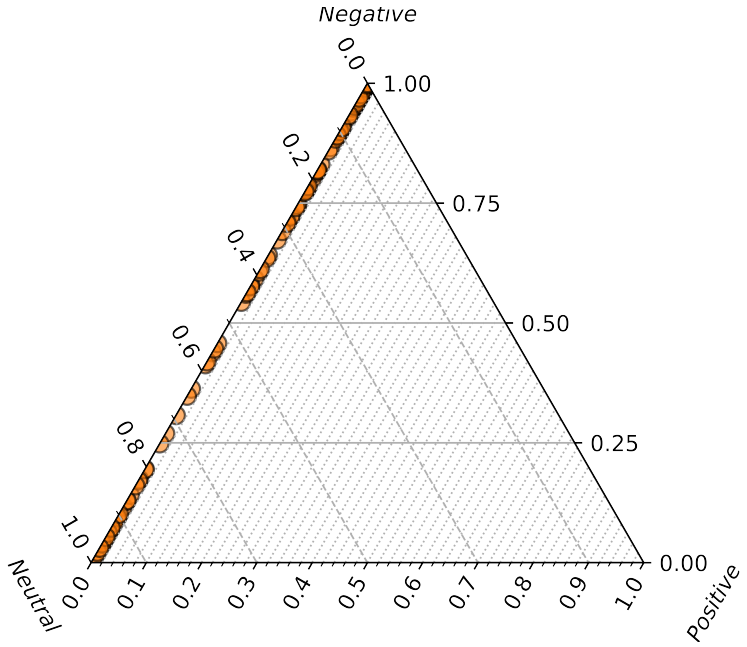


Figure 4.8: Rozloženie základných emócií v nadpisoch časti fake správy z datasetu FakeNewsNetpolitifact. Zdroj: autorský príspevok publikovaný v [29]

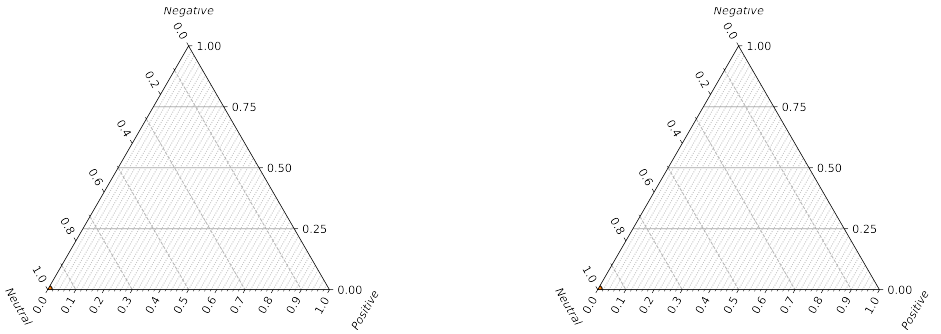
Po aplikovaní metódy citovo zafarbených slov na dataset FakeNewsNet [30] sme v prípade textov politickej časti dostali tieto výsledky. Pre texty označené autormi datasetu ako reálne správy vidíme na obrázku 4.7 na škále neutrálnych emócií a v prípade správ označených ako falošné na obrázku 4.8 vidíme všetky správy pomocou okometrickej metódy ako správy s negatívnym citovým zafarbením.

Ako vidno z grafov aj obyčajnou okometrickou metódou rozloženie základných emócií v grafe reálnych správ je prevažne v oblasti neutrálnej emócie, čo sa dá očakávať od nadpisov správ bez citového zafarbenia z politickej oblasti. Zároveň v grafe pre základné emócie falošných správ prevláda v nadpisoch týchto správ negatívna emócia. Týmto spôsobom sme si potvrdili našu hypotézu, že je možné zredukovať objem potrebného textu vstupujúceho do procesu detekcie a tým v podstatnej miere zvýšiť reálny počet prehľadávaných správ bez nutnosti pridávať ďalšie fyzické zariadenia s výpočtovým výkonom.

Podrobnejšie sa problematike zmenšovania objemu analyzovaného textu venujeme v našom pripravovanom článku, ktorý bude možné nájsť v [29] a toho času prechádza záverečnými úpravami.

V prípade anotácie dát z nami používaného vzorového datasetu sme využili základnú metódu evaluácie pomocou klasifikácie a detekcie citovo zafarbených slov (pozri sekcia 4.2.1). Každú správu sme ohodnotili na základe použitých výrazových prostriedkov na 3 zložky: negatívnu, neutrálnu a pozitívnu. Základný predpoklad je, že správy obsahujúce prvky antisociálneho správania budú mať vo väčšej miere zastúpené pozitívne a negatívne ohodnotené slová, ako v prípade reálneho textu. Tento predpoklad vychádza z pozorovania, že falošné správy útočia v prevažnej miere na emocionálnu zložku osobnosti prijímateľa správy.

Z výsledkov našej analýzy nám vyplýva, že v prípade správ z analyzovaného datasetu v politickej oblasti sme overili správnosť klasifikácie autorov datasetu.



(a) Reálne správy v časti FakeNewsNet datasetu v časti Gossip Cop podľa hodnôt základných emócií.

(b) Falošné správy v časti FakeNewsNet datasetu v časti Gossip Cop podľa hodnôt základných emócií.

Figure 4.9: Výsledky hodnôt zobrazené v trojstrannom grafe. Na obrázku 4.9a vidíme overenie klasifikácie v mediálnej oblasti na reálnych správach. Na obrázku 4.9b v prípade falošných správ.

Na rozdiel od overenia správnosti zatriedenia pomocou nadpisu v politickej oblasti vidíme v prípade mediálnej oblasti zobrazenej na obrázku 4.9 hodnotu v neutrálnej oblasti pri správach klasifikovaných ako nepravdivé, ale aj ako pravdivé. Na zobrazených výsledkoch vidno, kde dosahuje navrhnutá metóda predpokladané výsledky (v politických témach) a kde sme, naopak, narazili na limitácie metódy.

4.4.2 Sumarizácia

V nami navrhovanej metóde so skratkou Reduct_text_AI_LLM sme si overili jej možnosti na vstupných dátach zo sociálnej siete Twitter. Tematicky boli dáta zamerané na politickú oblasť a v druhej časti na oblasť médií. V oboch prípadoch bola metóda účinná v prípade pravdivých, ale aj nepravdivých správ. Pri správach z politickej oblasti však metóda dosahovala všeobecne lepšie výsledky. Podľa nášho úsudku sa lepšia detekcia dosahovala použitím metódy detekcie citovo zafarbených slov, ktorá sa vo väčšej miere prejavuje v prípade politických správ v podaní jednotlivých používateľov.

Rozdiel medzi mediálnou a politickou oblasťou si vysvetľujeme napríklad osobným zaujatím v hypotetickom prípade nič nerobenia štátnej zložky platenej z daní občanov, ako aj v prípade informácií o životnom osude známeho herca.

Výsledkom overovania pomocou citovo zafarbených slov v návrhu a evauluácie našej metódy Reduct_text_AI_LLM je úspešnosť limitne sa približujúca k 100% v politickej oblasti plnej emócií. Ak sa však overením emocionálnych slov pozrieme našou metódou na mediálnu oblasť, metóda zlyháva. Podľa nášho názoru nie je problém v samotnej metóde, ale vo verifikácii výsledkov pomocou emocionálne podfarbených slov.

Overenie novej metódy Reduct_text_AI_LLM tak naznačuje budúcu prácu na overovaní bez ovplyvnenia emóciami použitými pri písaní analyzovaného textu.

5 Záver

Našou analýzou sme získali ucelený pohľad na rôzne spôsoby antisociálneho správania používateľov sociálnych sietí v online priestore. Zoznámili sme sa s jeho formami a základnými charakteristikami, rovnako aj so spôsobmi ich riešenia podľa jednotlivých autorov. Prišli sme k záveru, že ohrozenou skupinou je mládež vo veku 15 až 18 rokov najmä z dôvodu nedostatku relevantných informácií a nemožnosti porovnať dané tvrdenie s prežitými udalosťami. Hranica takto ohrozenej skupiny sa posúva rovnako, ako sa posúva získanie zážitkov v reálnom svete.

Zo zistených skutočností vyplýva, že aj uznávané a mienkotvorné média mávajú problémy s nepravdivými informáciami. V niektorých prípadoch možno zachytiť 40% až 50% nepravdivých alebo skreslených informácií, ktoré sa dostanú do obsahu týchto médií aj cez redakčnú kontrolu (pre relevantné médiá tak vzniká povinnosť uverejniť opravu a ospravedlnenie). V prípade rôznych sociálnych fór je tento priestor pod nižšou mierou kontroly.

Takéto antisociálne správanie nie je fenoménom len nižšie vzdelaného obyvateľstva, ale vyskytuje sa v celom priereze spoločnosti. Zároveň sa aj autori rôznych publikácií dostávajú do situácií, kde majú možnosť vytvoriť falošné hodnotenia s cieľom dosiahnuť určitý zisk. Vhodným nástrojom na získanie čo najväčšieho množstva čitateľov je publikovanie informácií zameraných na citovú stránku, alebo na zvedavosť ľudí, ale najmä využívanie vhodných ustálených slovných spojení typu: “Unikla tajná informácia...”, “Dôveryhodné inštitúcie XY varujú...”, “Zdieľajte, pokiaľ to nezmažú...” a podobne.

Prínos našej predloženej práce je v tvorbe 4 nových metód, ktoré vzišli z našej analýzy danej problematiky. Našou snahou bolo vytvoriť nové metódy vytvoriť a zaradiť ich do správneho poradia podľa životného cyklu antisociálneho správania. Celkové zaradenie podľa publikácií je nasledovne:

- **Zber údajov** - prevažne v nami publikovanom článku [34] metódou na automatizované dlhodobé získavanie údajov zo zvolenej sociálnej siete.
- **Detekcia** - prevažne v nami publikovanom článku [29] a v [50].
- **Vznik antisociálneho správania a závislostí** - prevažne v nami publikovanom článku [2].
- **Redukcia potrebného objemu detekcie** - prevažne v nami publikovanom článku [29].
- **Forenzné vyšetrovanie** - prevažne v nami pripravovanom článku [43].
- **Ukotvenie do právneho rámca** - prevažne v nami pripravovanom článku [49].
- **Predikcia** - prevažne v nami publikovanom článku [2].

Zoznam bibliografických odkazov

[1] W. Pavlov and Skinner, "The history of psychology - behaviorism and humanism," *Introduction to Psychology*, 2023. [Online]. Available: https://courses.lumenlearning.com/waymaker-psychology/chapter/reading-behaviorism/

[2] I. Stupavský and P. Dakić, "Antisocial behavior and the dopamine loop on different technological platforms and industries: An overview," *Proceedings of Eighth International Congress on Information and Communication Technology (ICICT 2023)*, pp. 471–481, 2024.

[3] L. Selecká, I. Václavíková, M. Blatný, and M. Hrdlička, "Typológia antisociálneho správania: špecifiká prejavov adolescentných chlapcov a dievčat vo vzťahu k rizikovému sexuálnemu správaniu," *Česká a Slovenská Psychiatrie*, vol. 113, no. 6, pp. 258–267, 2017. [Online]. Available: http://www.cspsychiatr.cz/dwnld/CSP_2017_6_258_267.pdf

[4] S. Kumar, J. Cheng, and J. Leskovec, "Antisocial behavior on the web: Characterization and detection," in *WWW '17 Companion: Proceedings of the 26th International Conference on World Wide Web Companion*, Geneva, Switzerland, 2017, pp. 947–950. [Online]. Available: https://dl.acm.org/doi/10.1145/3041021.3051106

[5] P. Návrat, D. Chudá, M. Bieliková, M. Šimko, J. Šimko, I. Srba, M. Kompan, R. Móro, I. M. Ondik, P. Lacko, A. Martonová, K. Machová, J. Paralič, P. Butka, P. Bednár, M. Sarnovský, B. Mesárošová, R. Blaho, and L. Sabová, "Rebelion – odhaľovanie antisociálneho správania na webe," in *DATA A ZNALOSTI & WIKT 2018, Brno 11.-12.10.2018*, Brno, Česká republik, 2018, pp. 65–69. [Online]. Available: http://daz2018.fit.vutbr.cz/DaZ_WIKT_2018_Sbornik.pdf

[6] V. A. Achkasova, N. N. Zhuravleva, E. A. Trefilova, and L. V. Azarova, "Propaganda content in digital space: Identification markers," in *2019 Communication Strategies in Digital Society Workshop (ComSDS)*, 2019, pp. 88–92. [Online]. Available: https://ieeexplore.ieee.org/document/8709625

[7] V. Bahna, "Argumentum ad Hominem. Argumentation Strategies of Conspiracy Theories Advocates in Slovak Internet Discussions," *Národopisná revue (Journal of Ethnology)*, vol. 27, no. 4, pp. 298–306, 2017. [Online]. Available: https://www.academia.edu/35679028/Argumentum_ad_hominem_Argumenta

[8] A. Hrčková, I. Srba, R. Móro, R. Blaho, J. Šimko, P. Návrat, and M. Bieliková, "Unravelling the basic concepts and intents of misbehavior in post-truth society," *Bibliotecas. Anales de Investigación*, vol. 15, no. 3, pp. 421–428, 2019. [Online]. Available: https://www.researchgate.net/publication/333566984_Unravelling_the_basic_concepts_and_intents_of_misbehavior_in_post-truth_society

[9] M. Kosfeld, M. Haylock, P. KampkÃ¶tter, and F. Von Siemens, "Helping and Antisocial Behavior in the Workplace," no. 18154, May 2023. [Online]. Available: https://ideas.repec.org/p/cpr/ceprdp/18154.html

[10] L. Gangadharan, P. J. Grossman, and J. Vecci, "Antisocial behavior in the workplace," pp. 1–26, 2020. [Online]. Available: https://doi.org/10.1007/978-3-319-57365-6_139-1

[11] E. Ruijer, "Designing and implementing data collaboratives: A governance perspective," *Government Information Quarterly*, vol. 38, no. 4, p. 101612, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0740624X21000484

[12] P. Vivekananth, "The role of social media forensics in digital forensics," *International Journal Of Engineering And Management Research*, vol. 12, pp. 1–3, 2022.

[13] H. Farid, "An overview of perceptual hashing," *Journal of Online Trust and Safety*, vol. 1, no. 1, Oct. 2021. [Online]. Available: https://tsjournal.org/index.php/jots/article/view/24

[14] A. Shere, "Reading the investigators their rights: A review of literature on the general data protection regulation and open-source intelligence gathering and analysis," *The New Collection*, vol. 3, pp. 3–21, 2020.

[15] S. Yu, "Cyber profiling: Predicting political orientation with socmint," *Telematics and Informatics Reports*, vol. 10, p. 100058, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S277250302300018X

[16] A. Griffin, "Super-private social network launched to take on facebook with support of anonymous," https://www.independent.co.uk/tech/superprivate-social-network-launched-to-take-on-facebook-with-support-of-anonymous-10325307.html, Jun. 2015, accessed: 2023-07-16.

[17] I. U. Goicuría, A. G. Martín, and M. T. L. Gracia, "Inteligencia geospacial: Geoint," *Armas y Cuerpos*, no. 147, pp. 29–37, 2021.

[18] nga.mil, "Know the world, show the way. . . from seabed to space," https://www.nga.mil/about/About_Us.html, 2021, accessed: 2023-07-16.

[19] G. Serscikov, "Humint operations abroad: Challenges to japan's intelligence services," *International Journal of Intelligence and CounterIntelligence*, vol. 0, no. 0, pp. 1–31, 2023. [Online]. Available: https://doi.org/10.1080/08850607.2023.2214325

[20] R. P. Dwivedi, "The human intelligence: Creativity, communication and culture in crises," *International Journal of Culture & Global Studies*, vol. 4, no. 2, pp. 558–561, 2023. [Online]. Available: https://www.researchgate.net/profile/Ram-Dwivedi/publication/371279971_The_Human_Intelligence_Creativity_Communication_and_Culture_in_Crises/links/64a0d529c41fb852dd448a0c/The-Human-Intelligence-Creativity-Communication-and-Culture-in-Crises.pdf

[21] E. D. Liddy, "Enhanced text retrieval using natural language processing," *Bulletin of the American Society for Information Science and Technology*, vol. 24, no. 4, pp. 14–16, 1998.

[22] J. Staš, D. Hládek, S. Ondáš *et al.*, "Spracovanie prirodzeného jazyka pre interaktívne rečové rozhrania v slovenčine," *Proceedings ITAT 2015*, pp. 81–87, 2015.

[23] J. Paralič, K. Furdík, G. Tutoky, P. Bednár, M. Sarnovský, P. Butka, and F. Babič, *Dolovanie znalosti z textov*. Technická univerzita v Košiciach, 2010. [Online]. Available: https://jan.paralic.website.tuke.sk/knihy/DolovanieZnalostizTextov.pdf

[24] J. Schwarz, *Současný stav a trendy automatické indexace dokumentů: Propojení analytických záznamů s plnými texty*. Schwarz, 2003. [Online]. Available: http://full.nkp.cz/nkdb/docs/studie/MAIobsah.html

[25] M. Malỳ, "Natural language processing with application to slovak language," *Comenius University, BratislavaFaculty of Mathematics, Physics andInformaticsDepartment of Applied Informatics*, 2011 2011.

[26] S. Jaf and C. Calder, "Deep learning for natural language parsing," *IEEE Access*, vol. 7, pp. 131 363–131 373, 2019.

[27] M. Hu, "Cambridge analytica's black box," *Big Data & Society*, vol. 7, no. 2, p. 2053951720938091, 2020. [Online]. Available: https://doi.org/10.1177/2053951720938091

[28] J. Hinds, E. J. Williams, and A. N. Joinson, ""it wouldn't happen to me": Privacy concerns and perspectives following the cambridge analytica scandal," *International Journal of Human-Computer Studies*, vol. 143, p. 102498, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581920301002

[29] I. Stupavský, P. Dakić, and V. Vranić, "The impact of fake news on traveling and antisocial behavior in online communities: Overview," *Applied Sciences*, vol. 13, no. 21, p. 11719, Oct. 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/21/11719

[30] K. Shu, "Fakenewsnet," 2021. [Online]. Available: https://github.com/KaiDMML/FakeNewsNet

[31] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media," *arXiv preprint arXiv:1809.01286*, 2018.

[32] politifact, "Politifact - the poynter institute," https://www.politifact.com/, 2024.

[33] G. C. Media, "Gossipcop," 2024. [Online]. Available: http://www.gossipcop.com/

[34] K. Bielčiková, I. Stupavský, and P. Dakić, "Detection of misinformation using nlp methods and stylometry on social media in the slovak language," pp. 1–17, 2024.

[35] N. R. S. republiky (NRSR), "300/2005 z. z. Časová verzia predpisu účinná od 01.01.2020 do 31.10.2020 - § 140b trestné činy terorizmu," https://www.iso.org/iso-8601-date-and-time-format.html, 2024.

[36] T. Wang, G. Wang, B. Wang, D. Sambasivan, Z. Zhang, X. Li, H. Zheng, and B. Y. Zhao, "Value and misinformation in collaborative investing platforms," *ACM Trans. Web*, vol. 11, no. 2, May 2017. [Online]. Available: https://doi.org/10.1145/3027487

[37] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," *cornell.edu*, 2017.

[38] K. Shu, S. Wang, and H. Liu, "Beyond News Contents: The Role of Social Context for Fake News Detection," *cornell.edu*, 2017.

[39] S. Kai, "Fakenewsnet," https://github.com/KaiDMML/FakeNewsNet, 2021, accessed: 2019-01-15.

[40] i. twitter, "Twitter," https://twitter.com/, Jan. 2019, accessed: 2019-01-15.

[41] J. Soma, "Nrc emotional lexicon," http://jonathansoma.com/lede/algorithms-2017/classes/more-text-analysis/nrc-emotional-lexicon/, 2017, accessed: 2019-01-15.

[42] I. Stupavský and D. Chudá, "Prieskumná analýza prezidentských volieb 2016," *Data a znalosti a WIKT 2019, zborník konferencie*, pp. 162–166, 2019. [Online]. Available: https://hi.kkui.fei.tuke.sk/daz2019/DaZ_WIKT_2019_Zbornik.pdf

[43] I. Stupavský, R. Fajd, and P. Dakić, "The role of social networks in the detection of lone wolves attacks with potentially dangerous behaviors," *Journal of Information and Telecommunication*, 2024.

[44] R. Fajd, *Využitie digitálnej forenznej analýzy na sociálnych sieťach*. Fakulta informatiky a informačných technológií, Slovenská technická univerzita, 2023.

[45] ——, "Využitie digitálnej forenznej analýzy na sociálnych sieťach," https://opac.crzp.sk/?fn=detailBiblioForm&sid=2D2C61E4341E53665E62123BFA9A, 2023, accessed: 2023-07-09.

[46] N. R. S. republiky (NRSR), "Zákon č. 171/1993 z. z. zákon národnej rady slovenskej republiky o policajnom zbore," https://www.zakonypreludi.sk/zz/1993-171#hlava4, 2024.

[47] iso.org, "Iso 8601 date and time format," https://www.iso.org/iso-8601-date-and-time-format.html, 2024.

[48] G. riaditeľstvo pre komunikáciu Európskej komisie, "Akt o digitálnych službách," https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_sk, 2019.

[49] I. Stupavský, "Prediction and prevention of forms antisocial behavior in the legal environment of the slovak republic and the framework of the european commission," 2024.

[50] I. Stupavský and V. Vranić, "Analysing the controversial social media community," pp. 299–303, 2022.

A Publikácie autora

Zoznam publikačnej činnosti/List of publishing activities
Autor: STUPAVSKÝ, Igor

V3 Scientific output of publication activity in the journal

V3_01 DAKIĆ, Pavle - **STUPAVSKÝ, Igor** - TODOROVIĆ, Vladimir. The Effects of Global Market Changes on Automotive Manufacturing and Embedded Software. In Sustainability [elektronický zdroj]. Vol. 16, iss. 12 (2024), art. no. 4926, s. 1-31. ISSN 2071-1050 (2022: 3.9 - IF, Q2 - JCR Best Q, 0.664 - SJR, Q1 - SJR Best Q). V databáze: DOI: 10.3390/su16124926. Type of report: článok; Report: zahraničný; Publication category to 2021:ADE

V3_01 **STUPAVSKÝ, Igor** - DAKIĆ, Pavle - VRANIĆ, Valentino. The Impact of Fake News on Traveling and Antisocial Behavior in Online Communities: Overview. In Applied Sciences. Vol. 13, iss. 21 (2023), art. no. 11719, s. [1-21]. ISSN 2076-3417 (2022: 2.7 - IF, Q2 - JCR Best Q, 0.492 - SJR, Q2 - SJR Best Q). V databáze: WOS: 001100297000001 ; CC: 001100297000001 ; DOI: 10.3390/app132111719. Type of report: článok; Report: zahraničný; Publication category to 2021:ADC

V2 Scientific output of publication activity as part of an edited book or collection

V2_01 ALEKSIĆ, Marijana - DAKIĆ, Pavle - **STUPAVSKÝ, Igor** - TODOROVIĆ, Vladimir. Time and company management in cases of fake news within the automotive industry. In Economic and Social Development [Book of Proceedings] : 96th International Scientific Conference on Economic and Social Development - "Era of Global Crisis". 1. ed. Varazdin : Varazdin Development and Entrepreneurship Agency, 2023, S. 56-66. ISSN 1849-7535. Type of report: príspevok z podujatia; Report: zahraničný; Publication category to 2021:AFC

V2_02 ALEKSIĆ, Marijana - DAKIĆ, Pavle - **STUPAVSKÝ, Igor**. Design and implementation of quality management products and services. In Modern Challenges in Management, Economy, Law, Security and Information Society: Book of abstracts. 1. ed. Belgrade : MB University, 2024, S. 24. ISBN 978-86-6375-144-6. Type of report: abstrakt z podujatia; Report: zahraničný; Publication category to 2021:AFG

V2_03 DAKIĆ, Pavle - ALEKSIĆ, Marijana - **STUPAVSKÝ, Igor**. Evolution and significance of cloud computing: a comprehensive analysis in automotive business operations. In Modern Challenges in Management, Economy, Law, Security and Information Society: Book of abstracts. 1. ed. Belgrade : MB University, 2024, S. 27. ISBN 978-86-6375-144-6. Type of report: abstrakt z podujatia; Report: zahraničný; Publication category to 2021:AFG

V2_04 CHALÁS, Filip - **STUPAVSKÝ, Igor** - VRANIĆ, Valentino. Discussion Manipulation, Language and Domain Dependent Models: An Overview. In 2023 Zooming Innovation in Consumer Technologies Conference (ZINC). 1. ed. Danvers : Institute of Electrical and

Electronics Engineers, 2023, S. 136-141. ISBN 979-8-3503-4772-2. V databáze: SCOPUS: 2-s2.0-85166204785 ; DOI: 10.1109/ZINC58345.2023.10174128 ; IEEE: 10174128. Type of report: príspevok z podujatia; Report: zahraničný; Publication category to 2021:AFC

V2_05 **STUPAVSKÝ, Igor** - CHUDÁ, Daniela. Prieskumná analýza prezidentských volieb 2016. In Data a znalosti a WIKT 2019, zborník konferencie, Košice, KONGRES Hotel Roca, Slovensko, 10. - 11. október 2019. 1. vyd. Košice : Technická univerzita v Košiciach, 2019, S. 162-166. ISBN 978-80-553-3354-0. Publication category to 2021:AFD

V2_06 **STUPAVSKÝ, Igor** - VRANIĆ, Valentino. A study of media texts in the Slovak language. In 2022 IEEE Zooming Innovation in Consumer Technologies Conference (ZINC 2022). 1. ed. Danvers : Institute of Electrical and Electronics Engineers, 2022, S. 220-224. ISBN 978-1-6654-8375-9. V databáze: SCOPUS: 2-s2.0-85136435221 ; IEEE: 9840736 ; DOI: 10.1109/ZINC55034.2022.9840736. Type of report: príspevok z podujatia; Report: zahraničný; Publication category to 2021:AFC

V2_07 **STUPAVSKÝ, Igor** - VRANIĆ, Valentino. Analysing the controversial social media community. In Informatics 2022 : 2022 IEEE 16th International Scientific Conference on Informatics. Piscataway : Institute of Electrical and Electronics Engineers, 2022, S. 299-303. ISBN 979-8-3503-1034-4. V databáze: SCOPUS: 2-s2.0-85153346214 ; DOI: 10.1109/Informatics57926.2022.10083476. Type of report: príspevok z podujatia; Report: zahraničný; Publication category to 2021:AFD

V2_8 **STUPAVSKÝ, Igor** - DAKIĆ, Pavle - TODOROVIĆ, Vladimir - ALEKSIĆ, Marijana. Historical aspect and impact of fake news on business in various industries. In Economic and Social Development [Book of Proceedings] : 96th International Scientific Conference on Economic and Social Development - "Era of Global Crisis". 1. ed. Varazdin : Varazdin Development and Entrepreneurship Agency, 2023, S. 380-389. ISSN 1849-7535. Type of report: príspevok z podujatia; Report: zahraničný; Publication category to 2021:AFC

V2_9 **STUPAVSKÝ, Igor** - DAKIĆ, Pavle. Antisocial Behavior and the Dopamine Loop on Different Technological Platforms and Industries: An Overview. In Proceedings of Eighth International Congress on Information and Communication Technology ICICT 2023. 1. vyd. Singapore : Springer, 2023, S. 471-481. ISSN 2367-3370. ISBN 978-981-99-3235-1. V databáze: DOI: 10.1007/978-981-99-3236-8_37. Type of report: príspevok z podujatia; Report: zahraničný; Publication category to 2021:AFC

V2_10 TODOROVIĆ, Vladimir - DAKIĆ, Pavle - **STUPAVSKÝ, Igor** - ALEKSIĆ, Marijana. The Influence of fiscal effect on industrial growth and resilient computing for the automotive industry by considering the negative effects of world politics. In Geopolitical, geoeconomical and geofinancial challenges 2023 : thematic proceedings = Geopolitički, geoekonomski i geofinansijski izazovi 2023 : tematski zbornik. 1. ed. Beograd : Institut za finansije i makroekonomsku politiku, 2023, S. 46-59. ISBN 978-86-81262-03-0. Type of report: príspevok z podujatia; Report: zahraničný; Publication category to 2021:AFC

XXX Uncategorized

GLOB_01 **STUPAVSKÝ, Igor** - CHUDÁ, Daniela. Fake News Detection by Using Emotionally Coloured Words. In Workshop on Antisocial Behaviour Detection. 1. vyd. Bratislava : Slovenská technická univerzita, 2019, S. 1-2. Publication category to 2021:GII

Statistics: publication category

Statistics: category of feedback

V3	V3 Scientific output of publication activity in the journal	02
V2	V2 Scientific output of publication activity as part of an edited book or collection	10
XXX	Uncategorized	01
Sum		**13**

1	Citations in foreign publications, registered in the Web of Science citation indexes and the SCOPUS database	10
2	Citations in domestic publications, registered in the Web of Science citation indexes and the SCOPUS database	7
3	Citations in foreign publications not registered in citation indexes Google Scholar	10
Sum		**27**