Univerzita Hradec Králové
Fakulta informatiky a managementu
Katedra informačních technologií

Automatizovaná detekce falešných zpráv

Bakalářská práce

Autor: Matěj Habr
Studijní obor: Aplikovaná Informatika

Vedoucí práce: Ing. Martina Husáková, Ph.D.

Hradec Králové Duben 2023

Prohlášení:

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně a s použitím uvedené literatury.

V Hradci Králové dne 24.4.2023 Matěj Habr

Poděkování:

Děkuji vedoucí bakalářské práce Ing. Martině Husákové, Ph.D. za metodické vedení práce a konzultace během vypracovávání bakalářské práce.

Anotace

Tato bakalářská práce se zabývá problematikou automatizované detekce falešných zpráv. V práci je shrnuta problematika v aktuálním kontextu a krátce i historie vývoje falešných zpráv. Cílem práce je návrh a implementace systému, který by byl schopen automaticky rozpoznat falešné zprávy s vysokou přesností predikce. V této práci je využito několik metod strojového učení a analýzy textu. Z implementace vycházejí konkrétní výsledky a jsou diskutovány s možností dalšího vývoje modelu.

Annotation

Title: Automated detection of fake news

This Bachelor Thesis deals with the issue of automated detection of fake news. The work summarizes the issue in the current context and briefly also the history of the development of fake news. The aim of the work is to design and implement a system that would be able to automatically recognize fake news with high prediction accuracy. In this work, several methods of machine learning and text analysis are used. The results of the implementation are then presented and the results and possibilities for further development of the model are discussed.

Obsah

1 Úvod

Za poslední roky, nejen kvůli pandemii nemoci Covid-19, stouplo množství dezinformačních, nenávistných a manipulativních zpráv daleko více než kdy předtím. Další velký případ šíření falešných zpráv je aktuální krize na Ukrajině. Zde je velice obtížné jednotlivé zprávy rozeznat a kvalifikovat jejich pravost. Dalším faktorem ovlivňujícím dezinformační scénu jsou pravidelné politické volby. Tyto volby jsou velmi často tak vyostřené, že se společnost takřka rozděluje. Do určité míry je to způsobeno právě šířením falešných zpráv cílených nejen na strach, ale i na jiné lehce napadnutelné lidské vlastnosti. Společenská diskuse je možný způsob boje proti falešným zprávám. Jednou z možností je automatizovaná detekce za pomoci strojového učení. Avšak naučit počítač rozeznat význam jednotlivých slov nebo i celého textu, je velmi obtížné. Jak tedy naučit počítač alespoň do jisté míry porozumět zprávám, které jsou denně vkládány na internet? Cílem této práce je tedy navržení možnosti detekce falešných zpráv za pomoci strojového učení, a dalších metod tak, aby z práce vzešly co nejlepší výsledky a úspěšné porovnání.

1.1 Trend

Falešné zprávy byly, jsou a budou existovat neustále. Jejich šíření je v každé době regulováno jinými způsoby. V dnešní době je však jejich šíření téměř neregulováno, a to začíná přerůstat v nebezpečnou, avšak účinnou zbraň. Pokud jsou falešné zprávy dlouhodobě neregulované, je poté o to obtížnější čtenáře přesvědčit o jejich nepravosti. Čím déle jsou čtenáři s falešnými zprávami ve styku, tím spíše se tyto zprávy pro čtenáře stanou pravdivými a poté už je takřka nemožné kohokoliv přesvědčit o opaku, ať už jsou důkazy jakékoliv. Chceme-li navrhnout účinné řešení pro jejich detekci, je nejprve nutné si je přiblížit, vyobrazit jejich aktuální trend a specifikovat postup detekce a vyhodnocení.

1.2 Falešné zprávy

Falešnými zprávami jsou myšleny cíleně lživé nebo klamavé příspěvky stavěné na klamavém nebo jinak manipulativním základu. Například Paul Watzlawick v knize „Jak skutečná je skutečnost?: mylné představy, klamání, porozumění.“ [1] argumentuje, že falešné zprávy bývají často výsledkem našich mylných představ a toho, jak interpretujeme a chápeme informace, které přijímáme. Naproti tomu Jiří Táborský popisuje falešné zprávy jako nástroj pro manipulaci s lidmi. To podkládá vědeckými studiemi a experimenty, které byly v tomto oboru prováděny. [2] Největším problémem falešných zpráv dnešní doby je jejich rychlé šíření, a to zejména mezi uživateli sociálních sítí a emailových klientů, čímž se může ovlivnit veřejné mínění. Je velice obtížné určit, jak moc velký vliv falešné zprávy mají, ale lze soudit, že celkem velký. Pokud by vliv neměly, pravděpodobně by je nikdo již nevyužíval a nemusel by se tento problém vůbec řešit. Je ale velmi obtížné falešné zprávy přímo vyvrátit, není-li dostatek informací a důkazů. Při vyvracení falešných zpráv je ale i stejným problémem čtenáře seznámit s důkazy, vedoucí k zodpovězení otázky proč je zpráva hodnocena jako falešná.

1.3 Aktuální kontext

Na základě výroční zprávy Bezpečnostní informační služby ČR z roku 2020 [3] se během pandemie zvýšila vlna negativních reakcí vůči zavedeným opatřením. BIS přisuzuje tuto skutečnost dezinformátorům, kteří se podíleli i na pořádání protestních akcí.

Nejsdílenější článek na platformě Facebook z roku 2019 dle průzkumu portálu businessinsider.com [4] byl: *„Trumpův dědeček byl pasák a daňový podvodník, jeho otec členem KKK“*.

Nejsdílenější článek v ČR na základě průzkumu portálu investigace.cz [5] byl v roce 2021: *„Soňa Peková potvrdila, že někdo vyrábí umělé laboratorní kmeny SARS-Cov-X a vypouští je obrazně řečeno jako z jeskyně, kterou je potřeba najít a zavřít! Biolož ka znovu potvrdila, že virus se nechová jako přírodní izolát a nejnovější mutace z Velké Británie“.*

Před vznikem internetu, a tedy možností tvořit dezinformace daleko rychleji a efektivněji, lze brát jako jednu z největších dezinformačních kauz tu s tabákovými firmami v USA. Jiří Táborský ve své knize [2] zmiňuje, že během 50. let dvacátého století tabákové firmy dokázaly dlouho tvrdit, že je kouření zdravé. A to i ve chvíli, kdy se začaly objevovat studie, které tvrdily opak. Tabákové firmy zvolily, i přes to, že věděly, jaký mají tabákové výrobky vliv na lidské tělo, tvrdý boj, a to všemi prostředky. Mezi nimiž jsou hlavně dezinformace. Tabákové firmy si dokonce začaly vyrábět vlastní studie. Dalo by se tedy říct, že zvolily vidinu zisku před lidským zdravím, a to je vysoká cena.

Pokusíme-li se nalézt nějaký počátek dezinformací na webu, zásadním milníkem je označován dle autorů Haigha T., Haigha M. a Kozaka [6] začátek ukrajinské krize v roce 2015, kdy vznikaly první větší dezinformační weby, které začaly masivně manipulovat či dezinterpretovat dění, čímž mohly sehrát svou roli v dalším dění. Dalším milníkem byla migrační krize do Evropy, a s ní nárůst agrese a nevole vůči nim. Dále pak dle Laudera nejistá kampaň a volby v USA, kde vyhrál Donald Trump. Nejaktuálněji pak pandemie Covid-19 a aktuální krize na Ukrajině.

Otázkou tedy může být, zda falešné zprávy mohou skutečně ovlivnit politicko-společenský vývoj. Tato otázka je podrobována zkoumání. Dle Allcotta & Gentzkowa [7] bylo prokázáno například to, že voliči Donalda Trumpa sdíleli několikanásobně více falešných zpráv než voliči Hillary Clinton.

Není prokázáno, kdo za falešnými zprávami stojí. Pokus investigativních novinářů poukázal na dva různé typy zdrojů. První typ dezinformačních webů je pravděpodobně tvořen pod vidinou zisku. Zpravodajské servery publikují takové zprávy, aby nalákaly nejvyšší možný počet čtenářů, a weby poté osadí velkým množstvím reklam. V tomto případě jsou zprávy z většiny neškodné a čtenář se z nich nedozví nic nového. Zde je ale dost často využíván tzv. „clickbait“. Autoři záměrně publikují takový titulek, aby nalákal co nejvyšší počet čtenářů. Dost často titulek nemusí přímo interpretovat stejné informace jako zbytek zprávy. Příkladem

clickbaitu může být titulek českého týdeníku Dotyk.cz: *„Archeolog zjistil, kdy nastane skutečný konec světa podle mayského kalendáře. Je to překvapivé.“* (Malá 2023) [8]. V tomto případě se pravděpodobně nejedná o falešnou zprávu. Titulek je vykonstruován tak, aby na něj klikl co nejvyšší počet čtenářů. Čtenář z titulku získává pocit, že se dozví nezvratné nebo až neuvěřitelné informace, ačkoliv to tak mnohdy není. Druhým typem je ovlivňování politického dění v zemích. Podle Kudrny [9] vedou stopy zejména v Evropě hlavně do Ruska. Jejich působení je patrné i v dalších zemích Evropy. Pro příklad Russia Today nebo Sputnik jsou zpravodajské servery s původem a správou z Ruska. A v dalších případech nasazování lidí do různých diskusí v masivním měřítku, čímž diskuse zaplavují ve prospěch svých zájmů.

Ačkoliv neexistují přímé důkazy působení Ruska v oblasti dezinformací, Bezpečnostní informační služba ČR na to v jejich výroční zprávě z roku 2020 upozorňuje:

> *„V oblasti ruských vlivových aktivit zůstalo i v loňském roce prioritou prosazování ruských zájmů prezentovaných sice jako zájem český, nicméně skutečné české zájmy poškozujících. V globálním kontextu průběžně sílil paradox ruských vlivových operací, kdy narativ směrem k ruskému publiku hlásal, že Rusko je obklopeno nepřáteli, zatímco narativ určený k českému (resp. globálnímu) publiku zdůrazňoval, že Rusko je naše spása.“* (BIS, 2020) [3]

V roce 2022, když začala krize na Ukrajině, bylo několik zpravodajských serverů státem zablokováno, jelikož se šířila opravdu velká vlna dezinformací. Toto ovšem moc nepomohlo, a bylo spíše kontraproduktivní. Dezinformátoři přišli na to, jak blokaci obejít, proto by bylo potřeba vytvořit nástroj, který dokáže obsah zanalyzovat během chvilky, kdykoliv a kdekoliv.

2 Možnosti detekce falešných zpráv, přístup k datům

I když se proti falešným zprávám bojuje všemožnými způsoby jako je např. vzdělávání, prevence nebo i seriózní média, nedaří se vůči závažnosti problému dostatečně bránit. A proto je dost pravděpodobně potřeba zasáhnout daleko tvrdší zbraní jako je například automatizovaná detekce zpráv a případné označení takových zpráv. V první řadě je potřeba dostat možnost ověřování faktů k lidem, kteří si fakta sami od sebe nedovedou ověřit [10]. Je tedy potřeba tyto zprávy ověřovat (i kdyby jen částečně) automatizovaně, už jen z hlediska jejich počtu je velmi nereálné je ověřovat manuálně. Pro úspěšný boj je třeba zprávy ověřovat i dostatečně rychle, nejlépe ve fázi, kdy se ještě nestihly rozšířit. Podle Matthew Liebermana již nebude mít vyvrácení zprávy žádný vliv, neudělá-li se to dostatečně včas [11].

Podle výzkumníků je potřeba zvolit dvě cesty, jednak cestu osvěty a vzdělávání a zároveň vyvinout automatizované nástroje, které pomohou všem při boji proti falešným zprávám [12]. Spoustu společností se již několik let ohrazuje a vynakládá snahu proti falešným zprávám bojovat. Facebook do boje dal nemalé peníze, avšak dezinformátoři si stále hledají kreativnější způsoby, jak ochrany obcházet, a tak snaha společností moc není vidět. Dalším problémem sociálních sítí je to, že potřebují vykazovat zisky a dost často na to používají metody, které jsou často považovány až za nemorální. Sociální sítě pracují na principu udržení pozornosti uživatele co nejdéle, aby tedy dosáhli co nejlepších výsledků. Mnohdy sahají i na choulostivá témata (mezi nimiž je třeba i politika) a uživateli tedy zobrazují i agresivnější témata, která obsahují zavádějící informace jen, aby je udrželi na vlastní sociální síti co nejdéle. Tímto se pak téma falešných zpráv na sociálních sítích dostává do konfliktu, protože pokud by s nimi sociální síť měla bojovat, způsobila by si tím vlastně škody a musela by vynaložit úsilí pro nalezení jiné cesty k ziskům. Zde se nabízí otázka, jak moc je tento boj s falešnými zprávami v kompetenci státních subjektů. Má-li mít nad tímto bojem nějakou moc stát, jak velkou moc, tak aby nedošlo k nechtěným efektům jako je například cenzura? Jak

moc mají být falešné zprávy regulované a jestli není ta správná cesta transparentnost, informovanost a vzdělávání?

2.1 Data

Pro úspěšnou detekci falešných zpráv za pomoci strojového učení je potřeba mít získaná ověřená data včetně jejich zdrojů a autorů. Daty se rozumí velké množství vyhodnocených jak už falešných zpráv, tak i těch nefalešných a relevantních.

K datům se lze dostat několika způsoby. Existují různé datasety (datové balíčky, obsahující spousty vyhodnocených záznamů), které lze získat v rámci výzkumu, či ke studiu volně na internetu. Ačkoliv jsou to mnohdy data adekvátní a ověřená, pro tuto práci byly vynechány, jelikož se nehodí do konstrukce systému pro automatizovanou detekci falešných zpráv. Systém bude získávat data v jiné (kratší) podobě, než je datasety obsahují, a proto je jejich využití nevhodné. Pro detekování falešných zpráv a jejich porovnávání byla využita možnost poslední.

Další možností je využití sběru dat, kdy dochází k pravidelným návštěvám zdrojů dat (zpravodajské servery, nebo jiné zdroje na základě specifické studie). Sběr dat lze provádět přímo, tzn. automat přejde na zdroj, získá data využitím „čtení“ textu neboli tzv. web-scrapping. Web-scrapping je technika pro kompletní stažení webové stránky a její plně automatické analýzy zdrojového kódu. Z takové analýzy jsou poté stažena data, jako například obsah článků, nebo jakékoliv jiné stránky. Tato technika může být mnohými weby odsuzována, nebo i přímo zakázána.

Jinou, a v této práci využitou možností u zpravodajských serverů, je využití RSS (Realy Simple Syndication; RDF Site Summary). Jedná se o strukturované články zpravodajských serverů ve formátu pro snadnější strojové čtení pro různé čtečky, zpravodajské kanály a další. Formát těchto článků je tvořen skrze značkovací jazyk XML, a obsahuje zpravidla titulek článku, krátký úryvek, datum zveřejnění, odkaz a další využitelné informace. Bohužel RSS mnohdy neobsahuje autora článků, a tak bude potřeba zvolit vhodný způsob, jak jej získat, nebo predikovat.

2.2 Možnosti detekce falešných zpráv

Možností pro detekci falešných zpráv je několik. Lze využít NLP (Natural Language Processing, česky zpracování přirozeného jazyka), pro zpracování velkého množství textových i netextových dat. NLP techniky se využívají pro „porozumění" obsahu textů, extrahování užitečných informací a dat jako například jména, citáty (prohlášení), odkazy, adresy a spousty dalších dat. NLP by mohlo být vhodné a využitelné při predikci autora zprávy. Například by model mohl dostat titulky a úryvky zpráv a jejich autorů, které by zanalyzoval a udělal si analýzu chování a stylu psaní autorů. Následně by podle tohoto způsobu predikoval autora nových dosud nevyhodnocených zpráv.

Strojové učení je ve výsledku dost rozsáhlé a lze ho využít několika způsoby. Jeden ze způsobů je základní použití, čímž je myšleno rozpoznání s datasetem falešných zpráv a postupné rozšiřování a učení dalších falešných zpráv. Strojové učení lze také využít pro hledání spojitostí mezi různými zprávami různých zdrojů. Jde tedy o porovnávání zpráv mezi sebou a vyhodnocovat jejich podobnost na základě textu a předchozích porovnání.

Samotné základní strojové učení není natolik spolehlivé, a je třeba pro klasifikaci zavádějících (falešných) zpráv využít více způsobů najednou a tyto detekce provázat např. skórem s využitím váhy podle specifikované důležitosti. U detekce zpráv by to mohlo být například využití zdroje, zprávy, jejího obsahu, autora a dalších možností. Provázání je třeba udělat takovým způsobem, kdy budou mít relevantní data vyšší důležitost, oproti datům, které mají nižší relevantnost. Stejně tak by měly mít přednost ta data, která jsou ověřená. A mezi těmito vahami udělat takový systém, který stanoví přesné skóre.

Detekce, která vyhodnocuje zprávy takovým způsobem, že jeho výstupem je pouze informace „pravdivá/falešná" nemůže nikdy plnohodnotně fungovat. Především proto, že lidem mnohdy takový způsob analýzy a vyhodnocení zpráv nestačí a většinou k tomu vyžadují transparentní informace, jak k tomuto výsledku systém

dospěl. Nejlépe s uvedením přesného postupu vyhodnocení, příznaků, které byly při vyhodnocování uváženy a brány v potaz. Nelze tedy nikomu tvrdit, že „Tato zpráva je falešná.“ z důvodu „Protože proto.“. Vhodné hodnocení by mohlo vypadat například následovně: „Tato zpráva je pravděpodobně falešná, a to z důvodů: nedostatečné, nebo žádné zdroje, titulek je clickbaitového typu, a zpráva obsahuje ten a ten zásadní argumentační faul, mějte se na pozoru!“. Takové vyhodnocení řekne o mnoho víc než to předchozí. Ale i tak to nemusí být dostačující a pro mnoho lidí nepřijatelné, nebo zavádějící. Vždy je potřeba myslet na to, že je lepší, aby taková analýza byla prováděna způsobem: „Tady jsou data proč by zpráva mohla být zavádějící, nebo relevantní.“. Oproti způsobu „Tady jsou data, proč je zpráva falešná, nebo pravdivá.“. Cílem automatizované detekce falešných zpráv je tedy vhodné primárně čtenáře informovat a dát mu nějaký nadhled nad aktuální čtenou zprávou.

Aby se tedy mohly zprávy automatizovaně detekovat, je potřeba navrhnout takový systém, který bude provádět analýzu s co nejvyšší přesností a transparentností.

3 Návrh vlastního systému

Systém byl navržen s cílem sběru dat tak, aby nedošlo k žádnému narušení přímého autorského díla autorů zpráv. Získávány jsou pouze titulky a části zpráv, pokud to situace umožňuje, je u každé zprávy uveden zdroj i autor. Systém je koncipován tak, aby byl zásah osobního úsudku co nejmenší, a vždy byla stanovena jasná pravidla klasifikace. I přes to nemůže být lidský faktor plně eliminován.

3.1 Manuální klasifikace falešných zpráv

Aby mohl automatizovaný systém začít fungovat, je třeba mít získaná základní data. K tomu je třeba mít správně vyhodnocená data. Zpráva je považována za zavádějící (falešnou) v případě, kdy je její manuálně vyhodnocené skóre pod 0 bodů. Každá zpráva má počáteční skóre, které je 0 bodů. Z tohoto skóre se poté na základě ruční kontroly odečítají body a to za: zavádějící nebo nedohledatelné zdroje, neodpovídající obrázky (tím se rozumí obrázky, které ke zprávě nepřipadají, jsou dohledatelné u obsahově jiných zpráv), zásadní argumentační fauly (viz 3.1.1) a neověřitelná fakta. Při manuální klasifikaci jsou taktéž zohledněny shodné zprávy u jiných zdrojů, které jsou u klasifikace k dispozici. U manuální klasifikace nehraje roli zdroj ani autor (ačkoliv je skóre obou zmíněných objektů k dispozici a ve finále je skóre této zprávy zahrnuto jak k autorovi, tak ke zdroji). Tím je hlavně myšleno to, že se při klasifikaci nehledí na preference osoby, která zprávu klasifikuje. Osoba taktéž své preference v hodnocení nesmí zohlednit, došlo by tak ke zkreslení a naučení této preference i automatizovanou klasifikaci. Aby byla možnost přičítat skóre i k autorovi, který není bohužel získáván plně automaticky, je potřeba při manuální klasifikaci doplnit autora ručně.

3.1.1 Argumentační fauly

> *„Argumentační fauly (neboli řečnické triky, argumentační klamy apod.) jsou v diskuzi používány za účelem přesvědčení oponenta či publika o správnosti tvrzení mluvčího bez ohledu na logickou platnost samotných argumentů. Mohou působit na emoce i na rozum, může se jednat o přímý útok, ale i o manipulativní vsuvky. Někdy se mohou dokonce jevit jako*

skvělé argumenty, nicméně podstata argumentační faulů spočívá v tom, že jejich logické závěry není možné aplikovat obecně.“ (Bezfaulu.net) [13]

Z této citace je celkem jasné, že se argumentační fauly dělí do několika kategorií. Mezi tyto kategorie patří: důraz na rozum, důraz na emoce, manipulativní obsah, chybná příčina, chybné vyvození a útok.

Důraz na rozum lze popsat apelem na naše zažité představy. Cílem je tedy čtenáři, nebo posluchači předložit taková fakta, která si lze podložit něčím např. zažitým, co už známe. Příkladem může být fráze: „Vždy to tak bylo, tak tomu tak musí být i teď.“.

Naproti tomu **důraz na emoce** si zakládá na potvrzování pravdy skrze emoce. Mnohdy je tento typ argumentačních faulů velmi účinný, jelikož je jeho typem lidská emoce, která není nijak náročně ovlivnitelná. Využívat se může například při nedostatku důkazů, kdy se poté lidé uchylují k využití emocí.

Manipulace obsahem je technika, při které se manipuluje předně se závěrem, který mají dokazovat. Mnohdy by to mohlo být označováno přímo za lži, ale nemusí tomu být vždy. Příkladem může být například situace, kdy si autor vybere jen ty důkazy, které mu napomohou k manipulativnímu závěru a přitom vynechá podstatný zbytek důkazů, který může utvářet závěr úplně jiný.

Chybná příčina je využívána v případech, kdy se z jevu A a jevu B vyvozuje souvislost, která nemá mezi jevy nutně žádnou reálnou souvislost. Pro příklad by se mohl uvést jev A: „V české republice se zvýšila míra dopravních nehod.“ a poté jev B: „Počet silničních staveb je rekordní.“. Z těchto dvou jevů by poté mohlo jít vyvodit „Jelikož se staví rekordní počet nových silničních staveb, zvýšila se míra dopravních nehod.“. Toto tvrzení nemusí být nutně pravdivé, ani nijak ověřené, a proto by to byl argumentační faul chybné příčiny.

Chybné vyvození je mírně podobný chybné příčině, ale v tuto chvíli se zaměřuje na důsledek. Příkladem chybného vyvození může být: „Z Ukrajiny utíkají jen ti bohatí, teď jsem nedávno viděl jet drahého mercedesa s ukrajinskou poznávací značkou.“. Tento příklad je ukázkou důkazu anekdotou, což je nejznámější z kategorie chybných vyvození.

Poslední kategorií je **útok**. Tato kategorie se objevuje skoro v každé politické diskusi. Cílem je zaměření se na osobu oponenta, a nikoliv jeho argumenty, které diskutuje. Může to být pro příklad „Nemáte pravdu, na to jste příliš mladý a ještě jste nic nezažil“.

Aby mohla být zpráva označena jakožto obsahující argumentační faul, bylo specifikováno, že musí být zásadní. Tím je myšleno to, že argumentační faul musí zprávu ovlivnit natolik, aby bylo jasně zřetelné, že byl ve zprávě použit právě argumentační faul.

3.2 Princip funkce systému

Funkce systému se dělí na několik fází a těmi jsou: základní sběr dat, základní zpracování (porovnání s již stávajícími daty), určení skóre a výsledná ruční kontrola. Aby tento model mohl vůbec fungovat je vyžadováno mít manuálně vyhodnocená data do začátku, na kterých může model poté začít stavět své výpočty a predikce. Navrhovaný systém tedy bez vstupních naučených dat postrádá smysl.

3.2.1 Hodnocení falešných zpráv

Jelikož pomocí strojového učení nelze jednoznačně označit zprávu nebo článek jako falešný nebo nepravdivý, bude v rámci navrhovaného systému přiřazováno zprávám skóre. Skóre (viz 3.1) určuje důvěryhodnost zprávy, tzn. vyšší skóre označuje vyšší důvěryhodnost.

Skóre zprvu ovlivňuje ověřitelnost (tzn. je-li zpráva interpretována více relevantními zdroji dodává to důvěryhodnosti), poté je ovlivněno i na základě skóre z předchozích zpráv zdroje, taktéž skóre ostatních zpráv vyhodnocených

jako shodné. Těmito čísly je poté utvořeno skóre pro zprávu a je poté odvozena důvěryhodnost zprávy. Skóre zprávy je sestaveno tedy zprvu skórem zdroje, které je bráno jako „odrazový můstek“ pro skóre, k tomuto skóre je poté zprůměrované skóre podobných zpráv. Skóre podobných zpráv je skládáno z procentuální podobnosti zprávy, predikovaného skóre (nebo případně již manuálního skóre, pokud byla podobná zpráva již vyhodnocena) a váhy, která se stanovuje na základě toho, zda-li má podobná zpráva skóre predikované, nebo již manuální. Kde manuální má váhu vyšší.

Skóre zdroje je utvořeno průměrem skóre zpráv tohoto zdroje. Skóre autora je tvořeno průměrem skóre jeho zpráv. Jak skóre zdroje, tak skóre autora upřednostňuje ty zprávy, které jsou již manuálně ověřené, tzn. tyto zprávy mají oproti ostatním vyšší váhu v celkovém zprůměrování.

Vytvořené skóre za pomoci automatického procesu není finální a je podrobeno manuální kontrole, která specifikuje přesné skóre. Následně jsou uchovávány skóre obě, pro pozdější kontrolu přesnosti predikcí.

3.2.2 Sběr dat

Systém má k dispozici manuálně specifikované zpravodajské servery, ze kterých pravidelně čerpá články a data. Pro sběr dat je zvolena metoda využívající RSS kanály, kde se získávají v pravidelných intervalech data (v aktuálním systému je to jednou za 60 minut). Pro tuto část byla zvolena knihovna „feed-extractor“. Knihovna získá data a převede je na univerzální objekt, který je poté snadno zpracovatelný. K tomuto zpracování je mu k dispozici pouze odkaz na RSS kanál zdroje jakožto argument. Z těchto dat jsou získávány: titulek, krátký popis a datum zveřejnění. Tato data jsou poté uložena do databáze pod odpovídající zdroj.

3.2.3 Diceův koeficient kostky

Pro základní zpracování zpráv je nutné znát funkci Diceova koeficientu kostky. Tato metoda využívá k porovnání s ostatními titulky početní obsah stejných dvojic sousedních znaků tzv. bigramy. Diceův koeficient se počítá jako 2krát počet prvků společných pro obě množiny děleno součtem počtu prvků v každé množině. Vzoreček pro výpočet podobnosti dvou vět vypadá následovně:

PODOBNOST = 2 * POČ. STEJNÝCH. BIGRAMŮ / (POČ. BIGRAMŮ VĚTY A + POČ. BIGRAMŮ VĚTY B)

Mějme tedy věty A a B, tedy: A = “Venku je pěkné počasí” a B = “Dnes je opravdu pěkné počasí”. Nejprve rozdělíme každou větu na bigramy.

Pro větu A to bude tedy:
“Ve”, “en”, “nk”, “ku”, "u “, " j”, “je”, "e “, " p”, “ěk”, “kn”, “né”, "é “, " p”, “po”, “oč”, “ča”, “as”, “sí”.

Pro větu B to bude:
“Dn”, “ne”, “es”, "s “, " j”, “je”, "e “, " o”, “op”, “pr”, “ra”, “av”, “vd”, “du”, "u “, " p”, “ěk”, “kn”, “né”, "é “, " p”, “po”, “oč”, “ča”, “as”, “sí”.

Počet bigramů společných pro obě množiny vět je 12 (jsou to: “je”,“e “,” p”,“ěk”,“kn”,“né”,“é “,” p”,“po”,“oč”,“ča”,“as”). Počet bigramů ve větě A je 19. Počet bigramů ve větě B je 26.

Diceův koeficient pro tyto dvě množiny se tedy vypočítá pro tyto věty následovně:

PODOBNOST = (2 * 12) / (19 + 26) = 24/45 ≈ 0.53

Z tohoto výpočtu je patrné, že je podobnost mezi těmito větami 0.53. (resp. 53%)

3.2.4 Základní zpracování

Po přijetí článků jsou nejprve nové zprávy porovnány s již získanými, maximálně 48 hodin starými zprávami s využitím strojového učení a Diceova koeficientu kostky. Poté jsou propojeny ty články, které mají podobně odpovídající titulky nebo obdobně odpovídající krátký popis, čímž se specifikují obsahem shodné zprávy z jiných zdrojů. Tato část dodá článku jistou důvěryhodnost v podobě kladného skóre, ale také je zohledněno již vytvořené skóre těchto zpráv. Tzn. je-li zpráva interpretována více relevantními (důvěryhodnými) médii, nabývá na důvěryhodnosti (je více pravděpodobné, že zpráva nebude podvrh, pakliže o ní píše více nezávislých autorů z různých zdrojů). Relevantním médiem se rozumí zdroj, jež má získané skóre nad hranicí 1 bodu. Pod touto hranicí jsou zdroje považovány za nespolehlivé, nebo zavádějící. Zprávy, které se nenachází u jiných zdrojů, nezískají žádné skóre v této fázi. V této fázi je taktéž predikován autor, jelikož není získáván automaticky, predikce probíhá na základě matematické metody TF-IDF (Term Frequency-Inverse Document Frequency). Tato metoda je využívána k určení důležitosti jednotlivých slov v dokumentu nebo textovém úryvku. Porovnání probíhá na základě již klasifikovaných zpráv autorů stejného zdroje, které jsou uskupeny jako korpus (elektronický soubor autentických textů, ve kterém je možné jednoduše vyhledávat jazykové jevy, z pravidla slovní spojení) zpráv. Z tohoto korpusu poté metoda porovnává relativní četnost výskytů slov v porovnávané zprávě. Na základě tohoto porovnání je poté vybrán ten autor, který má nejpodobnější styl psaní a je tedy výsledkem predikce.

3.2.5 Vyhodnocení skóre

Po vyhodnocení obsahu shodných zpráv, je u zprávy vyhodnoceno skóre. Vyhodnocení probíhá na základě skóre předchozích zpráv stejného autora a zdroje. Zde je provedena kontrola ve formě porovnání skóre autora a zdroje, má-li autor značně vyšší skóre než zdroj, pro který píše, předpokládá se, že je tento autor důvěryhodnější než celý zdroj. Resp. je zde předpoklad, že se autor nepodílí na zavádějících zprávách a je mu to při zprůměrování se zdrojem zohledněno. Toto skóre je poté spojeno s již předpřipraveným skórem z fáze porovnání s ostatními

články a je utvořeno finální skóre automatické detekce, které je poté zapsáno a předáno na manuální kontrolu.

3.2.6 Manuální kontrola

V poslední fázi je nepovinná manuální kontrola automatického vyhodnocení, kontrolu lze brát i jako částečné strojové učení. Manuální kontrolou lze skóre zprávy úplně přepsat, a to v případě, že automatické vyhodnocení nebylo provedeno úspěšně. Stejně tak lze při manuální kontrole oddělit ty zprávy, které nejsou ke kontrolované zprávě relevantní, čímž se provede celé automatické vyhodnocení znovu, již bez připojených zpráv. Zprávy využívané při vyhodnocování skóre, které jsou manuálně zkontrolované, mají vyšší váhu než zprávy dosud nezkontrolované. Z toho vyplývá, že jsou upřednostňovány předchozí zkušenosti a obzvláště ty, které má systém potvrzené „učitelem“. Na základě manuální kontroly je taktéž vypočítáváno procento úspěšnosti odhalování nedůvěryhodných (falešných) zpráv. Stejně tak je i zaznamenáván vývoj tohoto procenta v čase a v počtu zkontrolovaných zpráv pro měření výkonu a úspěšnosti systému. Při manuální kontrole je taktéž kontrolován predikovaný autor a v případě špatné predikce je k vyhodnocení i připsán správný autor.

3.3 Struktura systému

Aplikace je stavěna na skriptovacím jazyku JavaScript a nadstavbě NodeJS. Pro Diceův koeficient kostky je využita knihovna „string-similarity“. K uchování dat je zvolena databáze MariaDB a pro komunikaci s ní je využita knihovna „mysql2“. Systém je provozován na serveru, který má operační systém Windows Server 2019 Datacenter. Pro opakovaný cyklus byla zvolena knihovna „cron“, která zajišťuje pravidelné volání funkcí. Pro predikci autorů byla zvolena knihovna „natural“, přesněji její TF-IDF třída.

Obr. 1 - Ukázka struktury databáze, Zdroj: vlastní tvorba skrze dbdiagram.io

3.3.1 Struktura databáze

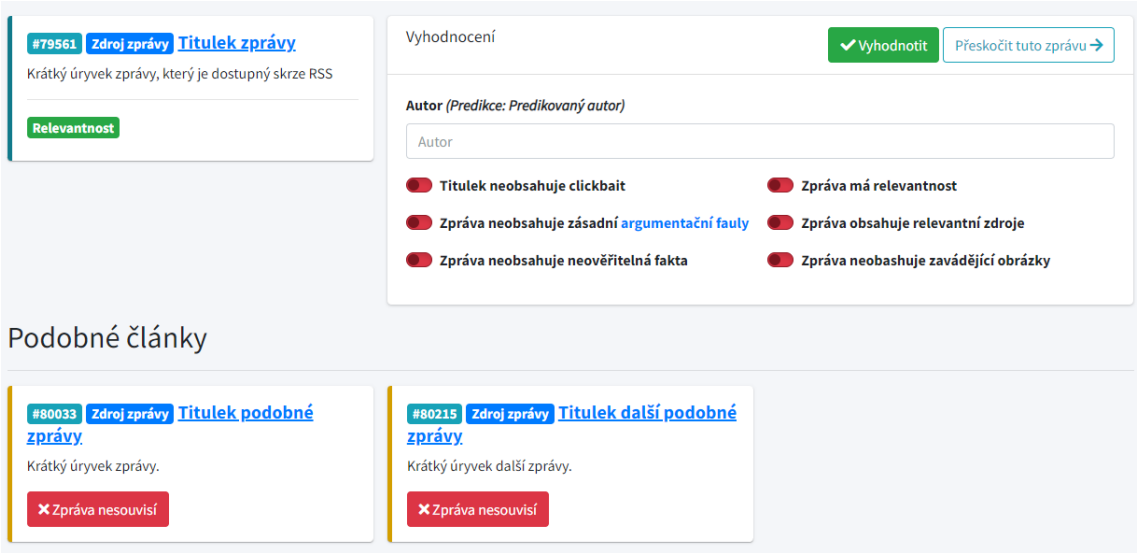
Hlavní tabulkou databáze jsou články, resp. tabulka „articles“, ve které jsou uchovávány veškeré zprávy. Každá zpráva má zdroj (source), titulek, krátký úryvek (body), datum publikování (published), odkaz, manuální skóre (výchozí 0), predikované skóre, predikovaného autora a datum manuální klasifikace. Tabulka zdrojů „sources“ obsahuje zdroje, z kterých jsou články získávány. Nachází se zde název, odkaz k RSS a skóre zdroje. Tato tabulka je taktéž spojována se zprávami a autory. Autoři jsou manuálně doplňováni při manuální klasifikaci. Každý autor má jméno, zdroj, pro který píše a skóre vypočtené na základě jeho zpráv. Jelikož může zpráva obsahovat více autorů, jsou autoři propojeni se zprávami skrze tabulku „articles_authors“, která slouží pouze k propojovacím účelům. Při vyhodnocení podobných zpráv jsou tyto možné propojení uváděny v tabulce „articles_links“ tato tabulka obsahuje id první a druhé zprávy, které jsou si podobné, skóre podobnosti, poté příznaky, zdali je toto propojení relevantní (obě zprávy pojednávají o stejném tématu). Poslední tabulkou je tabulka „article_values“, která zaznamenává hodnoty manuální klasifikace, veškeré příznaky, které může zpráva mít.

3.3.2 Rozvržená struktura

Hlavní část systému, zodpovědná za automatizovanou detekci, je stavěna pro neustálý běh. Systém je objektově orientovaný. Hlavní třída má cyklus opakovaný každých 60 minut, který získá a zpracuje nové zprávy pomocí instance objektu FeedAnalyzer. Dále má systém k dispozici třídu ArticleAnalyzer, která provádí vyhodnocení zpráv a následně je provedeno určení skóre. Komunikaci s databází obstarává třída Database stavěná na knihovně „mysql2“. V pravidelném intervalu je taktéž prováděna predikce autorů zpráv.

3.3.3 Struktura rozhraní

Pro rozhraní byl využit framework Svelte-Kit. Rozhraní je stavěno strukturálně bez objektů. Veškerá komunikace probíhá skrze API rozhraní. Rozhraní vyobrazuje pokaždé náhodně zvolenou nevyhodnocenou zprávu. K ní vyobrazuje seznam zpráv vyhodnocených jako podobných s možností je odebrat. Všechny zprávy lze skrze odkaz rozkliknout a přejít ke zdroji. U vybrané zprávy je formulář pro manuální kontrolu, zde se zaškrtávají ty příznaky, které zpráva splňuje. (viz. 3.1) Při potvrzení kontroly je náhodně vyobrazena další zpráva. U každé zprávy je taktéž nutné manuálně uvést jejího autora, k tomu je v rozhraní nachystané textové pole.



Obr. 2 - Ukázka vyhodnocení zprávy ve webovém rozhraní, Zdroj: vlastní tvorba

Webové rozhraní je dostupné na adrese https://fakenews.svelte.neoloop.cz. Na této adrese je možné provádět manuální klasifikaci zpráv. Taktéž je v rozhraní možnost ověřit predikci libovolné zprávy vložením titulku, úryvku, nebo i url adresy. Není povinnost vložit všechny tři parametry, s vyšším počtem parametrů je ale predikce skóre vyšší. Obsahuje-li databáze vloženou zprávu a je-li již manuálně ověřená, webové rozhraní nabízí její klasifikaci včetně již známých příznaků.

3.4 Hlavní funkce

Systém má mnoho funkcí a částí. Nejdůležitější z nich, které ovlivňují rozhodovací část systému, byly vyjmuty a uvedeny zde v práci. Mezi nejhlavnější funkce je pro příklad řazena funkce, která získává aktuální zprávy ze zdrojů. Vyhodnocovací funkce pro predikci skóre, nebo i autora. A rozhraní API pro komunikaci s webovým rozhraním.

3.4.1 FeedAnalyzer

Třída FeedAnalyzer obsahuje pouze jednu funkci a tou je analyze. Tato funkce získá a zpracuje do javascriptového objektu RSS objekt z adresy zdroje pomocí knihovny feed-extractor. Po dokončení extrakce zpráv je provedena automatická predikce skóre všech nových zpráv, taktéž je ověřena podobnost ostatních zpráv skrze třídu ArticleAnalyzer.

```
analyze(url) {
    return new Promise((resolve, reject) => {
        extract(url).then(data => {
            resolve(data);
        }).catch(err => reject(err));
    });
}
```

Ukázka 1 - Funkce analyze v objektu FeedAnalyzer, Zdroj: vlastní tvorba

Funkce analyze je využita při pravidelném získávání zpráv ze zdrojů. K tomu slouží funkce grabArticles. Funkce nejprve z databáze získá všechny zaznamenané zdroje, a poté každý postupně zkontroluje pomocí funkce analyze, která ze zdroje získá všechny aktuální zprávy. Ty jsou poté ověřeny, zdali už neexistují a pokud ne, jsou zaznamenány pro další analýzu.

```
function grabArticles() {
    Db.query(`SELECT *
              FROM sources`).then(res => {
        res.forEach(source => {
            FeedAnalyzer.analyze(source.url).then(articles =>
                articles.entries.forEach((article) => {
                    Db.query(`SELECT id
                              FROM articles
                              WHERE link = ?
                                AND source = ?`,
[article.link, source.id]).then(check => {

                        if (check.length === 0) {
                            let date =
article.published.includes("Z") ?
article.published.replace('T', ' ').replace('Z', '') : new
Date().toISOString().slice(0, 19).replace('T', ' ');
                            Db.query(`INSERT INTO articles
(source, title, body, published, link, predictedScore) VALUES
(?, ?, ?, ?, ?, ?);`, [source.id, article.title,
article.description, date, article.link, source.score]);
                        }
                    });
                });
            }).catch(err => {
              console.log(`${source.name} error ${err}`)
            });
        });
    });
}
```

Ukázka 2 - Funkce grabArticles pro získání nových zpráv, Zdroj: vlastní tvorba

3.4.2 Predikce autora

K predikci autora je využita funkce predictSource, která vyžaduje parametr sourceId, čímž je myšleno ID zdroje, přiřazené vytvořenou databází. Funkce získá všechny dosud manuálně neověřené články vybraného zdroje a provede predikci s využitím TF-IDF.

```
function predictSource(sourceId){
    const tfidf = new TfIdf();
    Db.query(`SELECT title,body,author_id FROM articles LEFT JOIN
articles_authors aa on articles.id = aa.article_id WHERE source = ? AND
aa.author_id IS NOT NULL`, [sourceId]).then(articles => {

        articles.forEach(article => {
            tfidf.addDocument(article.title + " - " + article.body,
article.author_id);

        });
        Db.query(`SELECT id,title,body FROM articles WHERE manualScore = 0
AND source = ?`, [sourceId]).then(newArticles => {
            newArticles.forEach(newArticle => {
                let maxMeasure = 0;
                let mostLikelyAuthor = null;
                let predctedAuthors = [];

                tfidf.tfidfs(newArticle.title + ' - ' + newArticle.body,
function(i, measure) {
                    if (measure > maxMeasure) {
                        maxMeasure = measure;
                        mostLikelyAuthor = tfidf.documents[i].__key;
                    }

                    let item = predctedAuthors.find(a => a.id ===
tfidf.documents[i].__key);
                    if(item){
                        item.measure += measure;
                        item.measure /= 2;
                    } else {
                        predctedAuthors.push({
                            id: tfidf.documents[i].__key,
                            measure: measure,
                        });
                    }
                });

                if(maxMeasure > 20) {
                    Db.execute(`UPDATE articles
                                SET predictedAuthor = ?
                                WHERE id = ?`, [mostLikelyAuthor === null ?
0 : mostLikelyAuthor, newArticle.id]);
                    console.log(`${newArticle.id} author is ${maxMeasure}
probably ${mostLikelyAuthor}!`);
                }
            });
        });
    });
}
```

Ukázka 3 - Funkce pro predikci autora zpráv, Zdroj: vlastní tvorba

3.4.3 ArticleAnalyzer

K predikci je využita třída ArticleAnalyzer, která má funkci predictScore, vstupními argumenty jsou všechny zprávy, všechna spojení mezi zprávami, všechny zdroje a přístup k instanci databáze z modulu mysql2. Funkce postupně projde každou zprávu, a pokud nemá ještě vyhodnocené manuální skóre, vypočte se na základě podobných zpráv a skóre zdroje predikované skóre. U podobných zpráv mají vyšší váhu ty zprávy, které už byly manuálně zkontrolovány. Skóre se počítá na základě vzorce: podobnost zprávy * skóre zprávy * váha. Takto vypočtené skóre se zprůměruje se všemi podobnými zprávami a skórem zdroje. A poté je predikované skóre zapsáno ke zprávě v databázi.

```
predictScores(articles, links, sources, Db){
    articles.forEach((article) => {
        let source = sources.find(src => src.id === article.source);

        if (parseFloat(article.manualScore) === 0) {
            let score = parseFloat(article.predictedScore);
            let similarLinks = links.filter(link => (link.firstArticle ===
article.id || link.secondArticle === article.id) &&
parseInt(link.predicted) === 0);

            if (similarLinks) {
                similarLinks.forEach((link) => {
                    let id = link.secondArticle !== article.id ?
link.secondArticle : link.firstArticle;
                    let similarArticle = articles.find(articl => articl.id
=== id);

                    let addScore = 0;
                    if(parseFloat(similarArticle.manualScore) === 0){
                        addScore =
(parseFloat(similarArticle.predictedScore)*(parseFloat(link.score))*0.2);
                    } else {
                        addScore =
(parseFloat(similarArticle.manualScore)*(parseFloat(link.score))*0.5);
                    }

                    score = (score + addScore) / 2;
                });
            }

            score = score !== 0 ? (score + parseFloat(source.score)) / 2 :
parseFloat(source.score);
            if(score !== parseFloat(article.predictedScore))
Db.query(`UPDATE articles SET predictedScore = ? WHERE id = ?`, [score,
article.id]);
        }
    });
}
```

Ukázka 4 - Funkce predictScores, která provádí predikci skóre, Zdroj: vlastní tvorba

ArticleAnalyzer má poté i funkci pro analýzu podobnosti zpráv. Podobnost se vypočítává na základě funkce z využité knihovny „string-similarity“, přesněji tedy konstanta stringSimilarity a její funkce findBestMatch, která zanalyzuje titulek s ostatními zprávami a z toho následně vrátí nejpodobnější titulky. Funkce analyzeArticles, která vypadá následovně:

```
analyzeArticles(articles, min = 0.4){

    let links = [];
    let analyzedTitles = [];
    let data = Array.from(articles, (a) => a.title);

    let f = true;
    articles.forEach((article) => {
        let title = article.title;

        if(!analyzedTitles.includes(title) && title.length < 300) {
            data.splice(data.indexOf(title), 1);
            let analyzed = stringSimilarity.findBestMatch(title, data);

            let relevantItems = analyzed.ratings.filter(item => item.rating
> min);

            if(relevantItems.length > 0){

                let toBeLinked = [];
                analyzedTitles.push(title);
                relevantItems.forEach((rItem) => {
                    analyzedTitles.push(rItem.target);
                    let foundArticle = articles.find(a =>
a.title.includes(rItem.target));
                    toBeLinked.push([foundArticle.id, rItem.rating]);
                });

                links.push({id: article.id, others: toBeLinked});
            }
        }
    });

    return links;
}
```

Ukázka 5 - Funkce analyzeArticles pro určení podobnosti zpráv, Zdroj: vlastní tvorba

Tato funkce se poté volá z funkce manageSimilarities(). Funkce získá z databáze všechny články napsané za poslední 2 dny a poté pomocí funkce analyzeAlrticles z třídy ArticleAnalyzer provede analýzu podobnosti zpráv. Podobné zprávy poté zapíše zpět do databáze, aby byla možnost toto spojení později využít při určení skóre.

```
function manageSimilarities() {
    Db.query(`SELECT id, title, published
              FROM articles
              WHERE published >= DATE_ADD(CURDATE(), INTERVAL -2
DAY);`).then(res => {
        if (res.length > 0) {
            let links = Articlator.analyzeArticles(res);
            links.forEach((link) => {
                link.others.forEach((other) => {
                    if (link.id !== other[0]) {
                        Db.query(`SELECT *
                                  FROM articles_links
                                  WHERE (firstArticle = ? AND secondArticle
= ?)
                                     OR (firstArticle = ? AND secondArticle
= ?)`, [link.id, other[0], other[0], link.id]).then(check => {
                            if (check.length === 0) {
                                Db.query(`INSERT INTO articleslinks
(firstArticle, secondArticle, score)
                                          VALUES (?, ?, ?)`, [link.id,
other[0], other[1]]);
                            }
                        });
                    }
                });
            });
        }
    });
}
```

Ukázka 6 - Funkce manageSimilarities, která analyzuje zprávy a jejich podobnost, Zdroj: vlastní tvorba

Jakožto relevantní podobnost mezi zprávami byla zvolena hranice 0.4 bodů, což odpovídá 40%. (klasifikace podobnosti je v rozmezí 0–1, kde 1 znamená „totožné“)

3.4.4 API

Další důležitou funkcí systému je API (Application Programming Interface), které je připravené pro práci s rozhraním. Funkčně odpovídá standardnímu REST API. Pro API byly navrženy tři end-pointy, které plní veškeré potřebné funkce k obsluze rozhraní.

3.4.4.1 End-point /unlink

Tento end-point slouží pro rozpojení propojení dvou zpráv, které spolu nesouvisí.

```
app.post("/unlink", (req, res) => {
    let body = req.body;
    Db.query(`UPDATE articles_links
              SET relevant = 0
              WHERE
          (firstArticle = ? AND secondArticle = ?) OR
          (firstArticle = ? AND secondArticle = ?)`,
          [body.firstArticle, body.secondArticle,
          body.secondArticle,
          body.firstArticle]).then(data => {
                  res.json({status: "ok"});
     });
});
```

Ukázka 7 - End-point /unlink, Zdroj: vlastní tvorba

3.4.4.2 End-point /article

Tento end-point navrací v JSON formátu náhodně zvolenou zprávu která ještě nebyla manuálně klasifikovaná. V JSON (JavaScript Object Notation) objektu je zpráva (titulek, krátký obsah, odkaz, zdroj, predikovaný autor). Pokud má zpráva nějaké podobné zprávy, tak jsou v JSON objektu i tyto všechny podobné zprávy včetně jejich zdrojů. Tento objekt je poté zpracován v rozhraní do uživatelsky přívětivé podoby.

```
app.get("/article", (req, res) => {
    Manager.getRandomUnseenArticle().then(article => {
        res.send(article);
    });
});
```

Ukázka 8 - End-point /article

```
getRandomUnseenArticle() {
    return new Promise((resolve, reject) => {
        Db.query(`SELECT articles.*,
                         av.*,
                         au.name as authorName,
                         sources.name  as source,
                         sources.score as sourceScore,
                         sources.id    as sourceId
                  FROM articles
                           JOIN sources ON source = sources.id
                           LEFT JOIN articles_values av on articles.id =
av.article
                           LEFT JOIN authors au on articles.predictedAuthor
= au.id
                  WHERE manualScore = 0
                  ORDER BY RAND()
              LIMIT 1`).then(res => {

                Db.query(`SELECT *
                      FROM articles_links
                      WHERE firstArticle = ?
                         OR secondArticle = ?`, [res[0].id,
res[0].id]).then(sims => {
                res[0].similarArticles = [];

                if (sims.length > 0) {
                    let ids = Array.from(sims, (a) => a.firstArticle !==
res[0].id ? a.firstArticle : a.secondArticle);

                    sims.forEach((sim) => {
                        Db.query(`SELECT articles.*,
                                         articles_links.score    as
similarityScore,
                                         articles_links.relevant as
relevantArticle,
                                         s.score                 as
sourceScore,
                                         s.name                  as source
                                  FROM articles
                                           JOIN articles_links
                                                ON firstArticle =
articles.id OR secondArticle = articles.id
                                           JOIN sources s on
articles.source = s.id
                                  WHERE (firstArticle IN (?)
                                      OR secondArticle IN (?))
                                    AND articles.id != ?
                                  GROUP BY articles.id`, [ids, ids,
res[0].id]).then(sime => {
                            res[0].similarArticles = sime;
                            resolve(res[0]);
                        });
                    });
                } else {
                    resolve(res[0]);
                }
            });
        });
    });
}
```

Ukázka 9 - Funkce getRandomUnseenArticle, Zdroj: vlastní tvorba

3.4.4.3 End-point /classify

Tento end-point očekává data o manuální klasifikaci zprávy, resp. její id, jednotlivé příznaky a autora. Všechna tato přijatá data jsou poté zahrnuta do databáze a je taktéž vypočteno a upraveno nové skóre zdroje a autora.

```
app.post("/classify", (req, res) => {
    let body = req.body;
    let authors = body.author.includes(",") ? body.author.split(",") :
[body.author];

    authors.forEach(author => {
        Db.query(`SELECT *
                  FROM authors
                  WHERE name LIKE ?
                    AND source = ?`, ["%" + author + "%",
body.source]).then((result) => {
            if (result.length > 0) {
                Db.query(`UPDATE authors
                          SET score = (score + ?) / 2
                          WHERE id = ?`, [body.score, result[0].id]);
                Db.query(`INSERT INTO articles_authors (article_id,
author_id)
                          VALUES (?, ?)`, [body.id, result[0].id]);
            } else {
                Db.query(`INSERT INTO authors (name, source, score)
                          VALUES (?, ?, ?);`, [author, body.source,
body.score]).then((inserted) => {
                    Db.query(`INSERT INTO articles_authors (article_id,
author_id)
                              VALUES (?, ?)`, [body.id,
inserted.insertId]);
                });
            }
        });
    });

    Db.query(`UPDATE articles
              SET manualScore = ?,
                  resolved    = NOW()
              WHERE id = ?`, [body.score, body.id]);
    Db.query(`UPDATE sources
              SET score = ?
              WHERE id = ?`, [body.sourceScore, body.source]);
    Db.query(`REPLACE INTO articles_values (article, clickbait, faults,
facts, relevant, sources, images)
              VALUES (?, ?, ?, ?, ?, ?,
                      ?);`, [body.id, body.clickbait, body.faults,
body.facts, body.relevant, body.sources, body.images]);
    res.json({status: "ok"});
});
```

Ukázka 10 - End-point /classify, Zdroj: vlastní tvorba

4 Shrnutí výsledků

Pro systém bylo během časového intervalu zhruba půl roku nasbíráno zhruba 80.000 různých zpráv z 47-mi různých zdrojů. Manuálně klasifikováno pro sledování vývoje přesnosti skóre a autorů bylo zhruba 1.200 náhodně vybraných z nich. Důvodem manuální klasifikace je hlavně slušný základ dat, na kterých lze pozorovat vývoj automatizované detekce. Tím lze říct, že čím více manuálně ověřených dat, tím je model více natrénován. Z tohoto počtu klasifikovaných zpráv byly poté utvořeny odpovídající grafy a přehledy jejich vývoje.

4.1 Zdroje zpráv

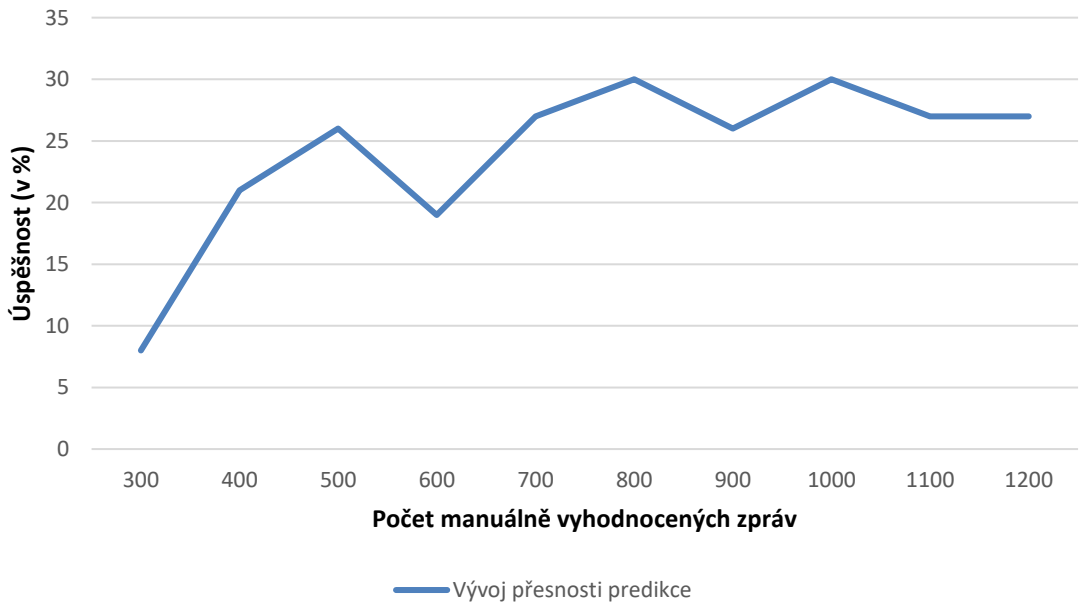
Do systému byly pro predikci namátkově přidány tyto servery (uváděny vč. url adresy):

- Novinky.cz (https://novinky.cz),
- ac24.cz (https://www.ac24.cz),
- AENews (https://aeronet.news),
- idnes.cz (https://idnes.cz),
- Seznam zprávy (https://www.seznamzpravy.cz),
- Parlamentní listy (https://www.parlamentnilisty.cz),
- České noviny (https://www.ceskenoviny.cz),
- TN.cz (https://tn.nova.cz),
- ČT 24 (https://ct24.ceskatelevize.cz),
- CNN Prima NEWS (https://cnn.iprima.cz),
- Aktuálně.cz (https://www.aktualne.cz),
- tadesco.org (https://tadesco.org),
- islamizace.cz (https://www.islamizace.cz),
- echo24.cz (https://echo24.cz),
- iRozhlas.cz (https://irozhlas.cz),
- Lidovky.cz (https://lidovky.cz),
- Hospodářské noviny (https://hn.cz),
- Křesadlo.com (https://kresadlo.com),
- RefleX (https://www.reflex.cz),

- Britské listy (https://www.blisty.cz),
- E15 (https://www.e15.cz),
- Týden.cz (https://www.tyden.cz),
- prahaIN (https://www.prahain.cz),
- pravda24 (https://pravda24.cz),
- newsbox.cz (https://newsbox.cz),
- CzechCrunch (https://cc.cz),
- Dotyk.cz (https://www.dotyk.cz),
- Pravý prostor (https://pravyprostor.net),
- Zvědavec (https://zvedavec.news),
- Pravdivě (https://pravdive.eu),
- VIP Noviny (https://www.vipnoviny.cz),
- Věk Světla (https://veksvetla.cz),
- Nejvíc info (https://www.nejvic-info.cz),
- Důležité24 (https://zpravy.dt24.cz),
- Otevři svou mysl (https://otevrisvoumysl.cz),
- RaptorTV (https://raptor-tv.cz),
- Zvedavec.news (https://zvedavec.news),
- Nová republika (https://www.novarepublika.online),
- Vaše věc (https://vasevec.parlamentnilisty.cz),
- Incorrect (https://www.incorrect.cz),
- Necenzurovaná pravda (https://necenzurovanapravda.cz),
- Aha! (https://www.ahaonline.cz),
- Blesk (https://www.blesk.cz),
- Antimeloun (http://www.antimeloun.cz),
- Česká věc (https://ceskavec.com),
- Kosa Nostra (https://vlkovobloguje.wordpress.com),
- Deník N (https://denikn.cz)

4.2 Výsledky predikce autorů

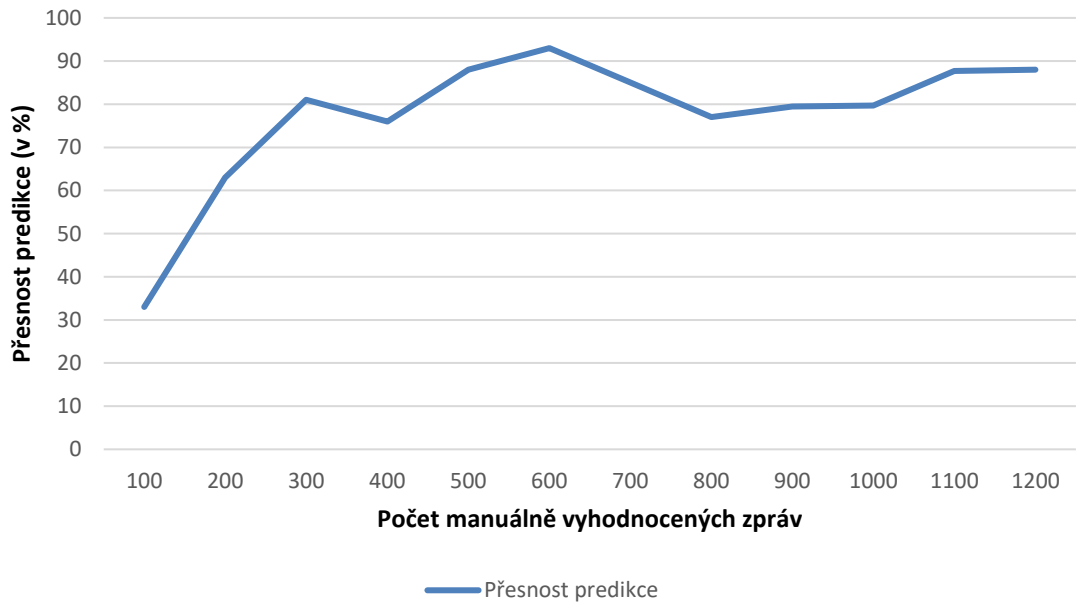
Manuální klasifikací se podařilo zvýšit přesnost predikce autorů až na 30%. Je zde ale nutné podotknout, že při manuální klasifikaci jsou zprávy vybírány náhodně a systém nemusí znát všechny autory daného zdroje. Dokud není autor manuálně přidán, systém ho nezná. Přesnost predikce je tedy vždy počítána od předchozí stovky vyhodnocených dat. Tzn. při ruční klasifikaci 500 zpráv je provedena predikce autorů (systém tedy zná autory 500 předchozích zpráv). Při dosažení další stovky je predikce provedena znovu, ale již obsahuje dalších 100 zpráv a autorů. A takto byla predikce prováděna vždy znovu na dalších datech. Pro kvalitní predikci je tedy potřeba daleko větší množství dat a znalost veškerých autorů a jejich zpráv. Je důležité podotknout, že predikci zkresluje ČTK (Česká tisková kancelář), které je přisuzováno autorství 31% manuálně zkontrolovaných zpráv (z 1200) ať už jsou přímo z jejich serverů, nebo i z jiných. Na 1.200 zpráv z různých zdrojů je zaznamenáno zhruba 500 autorů.



Obr. 3 - Graf vývoje přesnosti (v %) predikce autorů při manuální predikci, Zdroj: vlastní tvorba

4.3 Výsledky predikce skóre zpráv

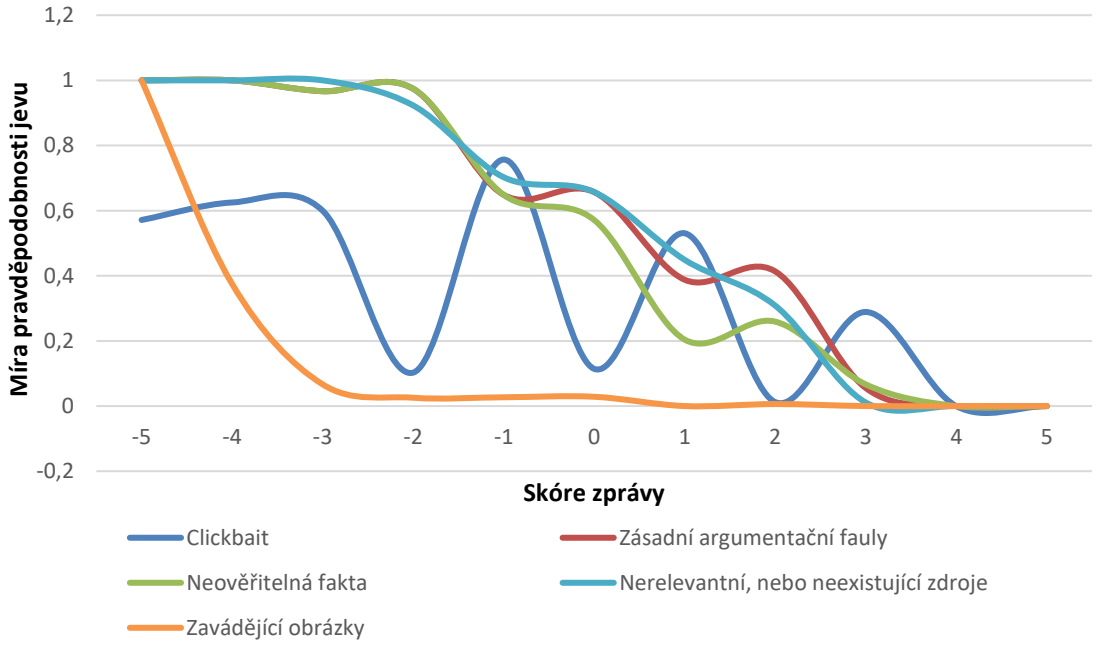
Přesnost je vypočítána pomocí rozdílu mezi manuálním skórem a predikovaným skórem. Predikované skóre se vždy po 100 manuálně zkontrolovaných zprávách přepočítává. Způsob přepočítání funguje na principu, kdy se z posledních 100 manuálně ověřených zpráv získá jejich manuální skóre, a to se porovná s predikovaným u těchto zpráv. Na základě průměru manuálních a predikovaných skóre se poté spočítá jejich průměr a tím je získána přesnost. Už při klasifikaci prvních 300 zpráv se přesnost vyšplhala nad 70%.



Obr. 4 - Graf vývoje přesnosti (v %) predikce skóre zpráv , Zdroj: vlastní tvorba

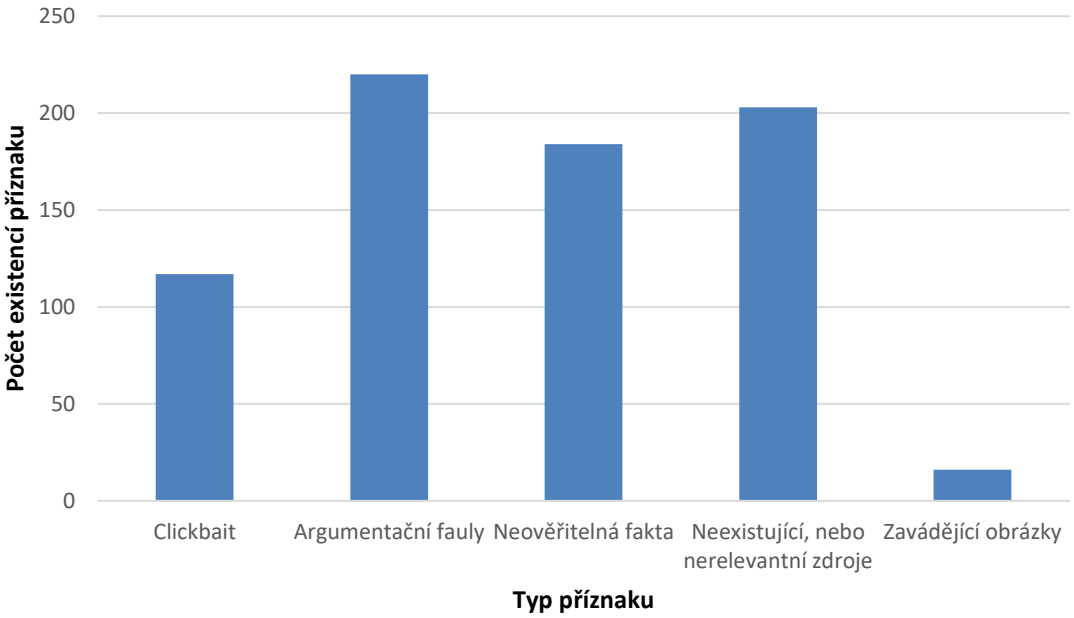
Lze tedy předpokládat, že se bude s narůstajícími daty přesnost postupně zvyšovat. Z aktuálních dat lze zjistit, že průměrné skóre manuálně klasifikovaných zpráv je **3.14** a průměrné skóre predikovaných dosud neklasifikovaných zpráv je **2.81**. Rozdílem mezi manuálním a predikovaným skórem je tedy **0.33** bodů. Lze tedy vyobrazit průměrný vývoj jednotlivých příznaků zpráv podle skóre. V grafu je maximum 1, což by mohlo mít význam např. „určitě obsahuje“ a minimum 0, které si lze vynaložit jako „neobsahuje“. Můžeme tedy sledovat, jak vypadají průměrné zprávy se skórem záporným, ale i kladným. Na základě těchto dat by byla možnost tyto příznaky predikovat. Z grafu lze vypozorovat, že zprávy se skórem pod -3

body budou dost pravděpodobně obsahovat zavádějící obrázek. Clickbaity se objevují v průběhu celého grafu od nejnižšího skóre až k 4 bodům, z toho lze předpokládat, že se objevují napříč zpravodajskými servery a jsou využívány jak u falešných zpráv, tak i u ostatních. Je to tedy celkem běžný prvek. Ostatní příznaky ubývají v plynulém spádu až ke skóre 4 bodů.



Obr. 5 – Graf změn příznaků podle skóre zprávy, Zdroj: vlastní tvorba

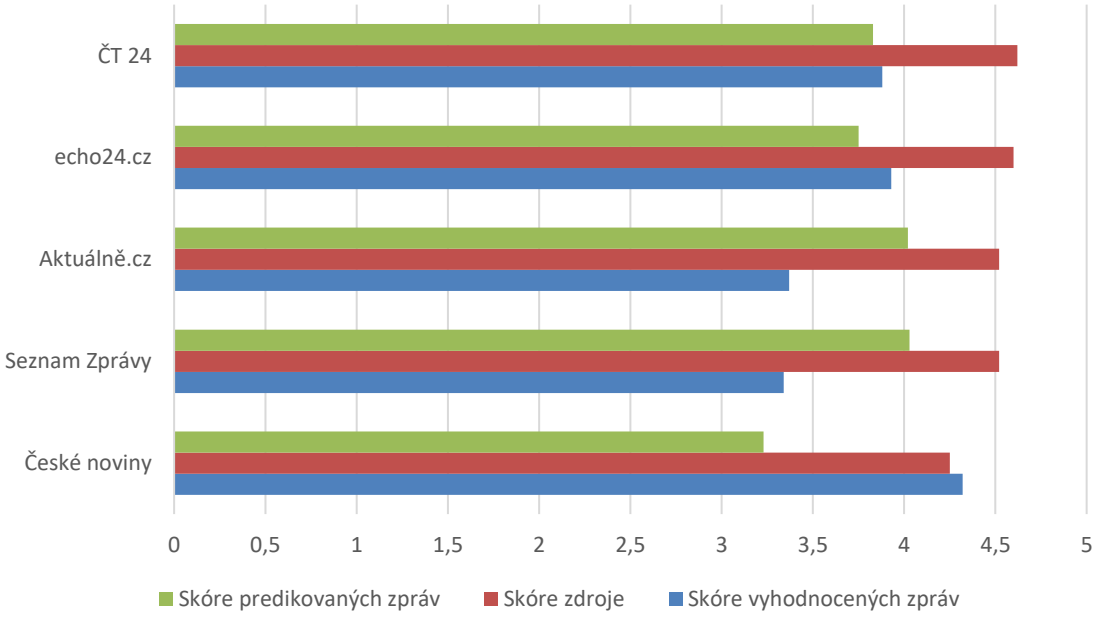
Data o příznacích si lze vyobrazit taktéž v celkovém počtu existencí při manuální klasifikaci 1.200 zpráv. Nejméně časté jsou zavádějící obrázky, které se objevují nejčastěji při skóre nižším než -3 body. Nejčastější jsou zásadní argumentační fauly, které se ve zprávách vyskytují celkem často. Je tedy možné, že jsou vkládány s nějakým úmyslem. Pro toto by mohlo být dobré vytvoření detekčního modelu na specifické argumentační fauly.



Obr. 6 – Graf počtu jednotlivých příznaků ve zprávách, Zdroj: vlastní tvorba

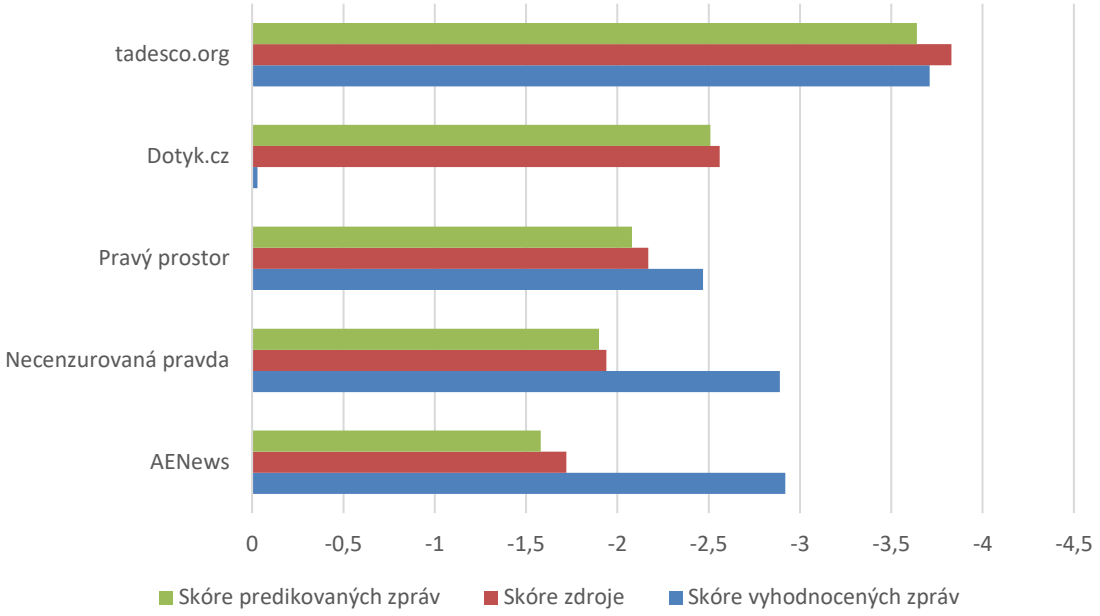
4.4 Výsledky predikce zdrojů

Při predikci skóre zdroje je vypočítáván průměr skóre zpráv. Na základě těchto dat můžeme vytvořit graf v němž můžeme porovnat průměrné skóre vyhodnocených zpráv, skóre zdroje a průměr predikovaného skóre dosud nevyhodnocených zpráv.



Obr. 7 - Graf predikce u 5 nejvýše hodnocených zdrojů, Zdroj: vlastní tvorba

Graf ukazuje nadhodnocení skóre 4 zdrojů z 5. Nejpodobnější skóre zdroje a vyhodnocených zpráv mají České noviny, které jsou řazeny až jako páté. Ve většině případů jsou všechny 3 skóre v rozsahu maximálně 1,5 bodu od sebe. Při vyšším množství manuálně ověřených zpráv by se očekávalo sblížení těchto skóre. Graf z druhé strany neboli od nejnižších skóre vykazuje následující údaje.



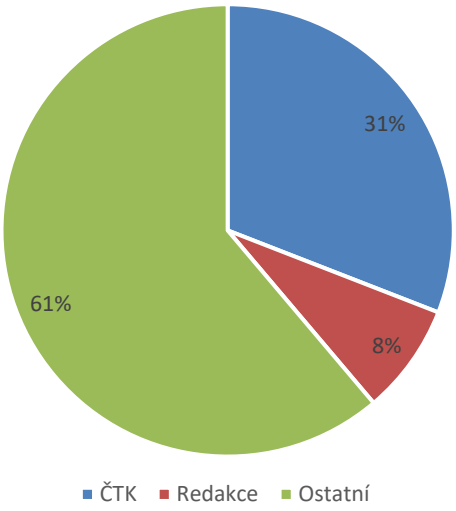
Obr. 8 - Graf predikce u 5 nejníže hodnocených zdrojů, Zdroj: vlastní tvorba

U záporných zdrojů jsou u převážné většiny predikované zprávy blíže ke skóre zdroje než-li ty vyhodnocené. Vysvětlení tohoto jevu by mohlo být jednoduché. Zpráv, kde je skóre nižší než 0 je početně méně, resp. i celkový počet zpráv ze zdrojů s nejnižším skóre je daleko méně než ostatních. Je tedy pravděpodobné, že se může manuální a predikované skóre rozcházet o něco více, než skóre u kladných zpráv. Bylo by tedy potřeba získat daleko více zpráv (jak manuálně ověřených, tak i predikovaných) z těchto nejníže hodnocených zdrojů, aby se manuální a predikované skóre více sblížilo.

4.5 Ostatní výsledky

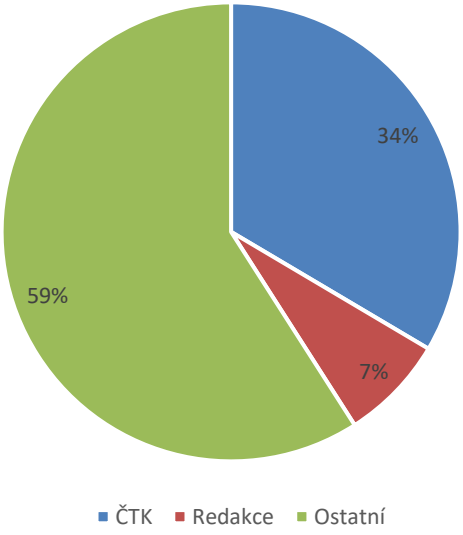
Během analýzy zpráv bylo taktéž zjištěno, že je u zhruba jedné třetiny zpráv přisuzováno autorství ČTK (České Tiskové Kanceláři). ČTK je institucí vydávající nejen zpravodajství, které následně mohou využívat ostatní zpravodajské servery.

Avšak při hlubší analýze těchto zpráv bylo zjištěno, že je mnohdy uveden ČTK jako jediný autor zprávy, ať už byla zpráva publikována na jakémkoliv zdroji. Též bylo zjištěno, že existují zpravodajské servery, které vydávají zprávy pouze s autorstvím ČTK a není u žádné jiné zprávy uveden jiný autor. Můžeme se tedy podívat na strukturu připisovaného autorství na základě manuální predikce.



Obr. 9 - Graf autorství zpráv podle manuální klasifikace, Zdroj: vlastní tvorba

Porovnání k tomuto grafu je predikované autorství u dosud nevyhodnocených zpráv, které vypadá následovně.



Obr. 10 - Graf autorství zpráv podle predikce, Zdroj: vlastní tvorba

Z grafů autorství je patrné, že predikce odpovídá v maximálním rozdílu 3% od manuálně přiřazených autorů. Mohlo by to tedy znamenat, že se predikce blíží skutečnosti.

Podíváme-li se na zdroje, nejnižší vypočtené skóre je **-3.8** bodů u tadesco.org, naproti tomu nejvyšší je **4.6** u ČT 24. Z 47 zdrojů má skóre nižší než 0 bodů 13 z nich. Kladné skóre má tedy 34 zdrojů.

Průměrné skóre všech dosud přidaných autorů je **3.05** bodů, což je více než nadprůměr škály, která se může pohybovat od -5 do 5. Průměrné skóre autorů je od průměrného skóre klasifikovaných zpráv vzdálené asi jen o 0.09 bodů, což je téměř neznatelný rozdíl a lze tedy předpokládat, že se budou tato skóre vyvíjet vzájemně podobně.

Pro statistický přehled lze specifikovat i poměr počtu autorů na počet zpráv. Z již získaných dat můžeme říci, že na 1.200 zpráv je přisuzováno jejich autorství celkem 541 autorům. (s tím, že mohou mít některé zprávy i více autorů, než pouze jednoho) Z 1.200 zpráv je dle předchozích grafů 31% přisuzováno ČTK, je tedy logické, že 828 zpráv nemá za autora ČTK. Naopak 372 zpráv má uvedeno jako autora ČTK.

5 Shrnutí

Prvním krokem k vytvoření modelu pro detekci falešných zpráv je studie literatury a souvisejících zdrojů. Dalším krokem bylo sestavení systému pro sběr zpráv z RSS zdrojů v pravidelných intervalech. Aby byla možnost ověření relevantnosti, bylo do systému sběru zpráv implementováno automatizované porovnávání podobnosti nových zpráv s využitím Diceova koeficientu kostky. Během sběru zpráv byl navržen model predikce falešných zpráv na základě podobnosti s ostatními zprávami.

Během získávání zpráv došlo k neočekávanému zádrhelu v podobě téměř stoprocentní absence autorů v RSS datech některých zdrojů. Aby byla možnost využít i autorovo skóre musel být model upraven tak, aby dovedl predikovat i autory. Pro predikci autorů byla využita matematická metoda TF-IDF, která dovede na základě již přiřazených zpráv k autorům vyhodnotit nejpravděpodobnějšího autora na základě skladby textu v titulku a úryvku zprávy. Již vyhodnocení autoři byli získáni během manuálních kontrol.

Aby mohl model začít predikovat bylo nutné zajistit již vyhodnocená data, bylo zkonstruováno webové rozhraní za pomocí frontend frameworku Svelte-Kit. Aby mohla fungovat manuální predikce bylo nutné přijít na základní příznaky falešné zprávy. Základem je tedy ustanovení takových hodnotitelných příznaků, které dopomohou k úspěšné a co nejtransparentnější detekci falešné zprávy. Na základě výzkumu byly zvoleny parametry: clickbaitový titulek, relevantnost, zásadní argumentační faul (viz 3.1.1), zavádějící obrázky a neexistující nebo nerelevantní zdroje. S těmito parametry a s co nejvyšší nezaujatostí započala manuální kontrola na náhodně volených zprávách.

Manuální kontrola byla provedena na 1200 zprávách. Po vyhodnocení každé sté zprávy byl model spuštěn pro predikci na dosud nevyhodnocených zprávách. Predikce se ke každé zprávě uložila (predikce je skóre zprávy a pravděpodobný

autor). Obdobně bylo při každé sté zprávě zapsáno z vyhodnocených zpráv manuální skóre a predikované skóre, které bylo uvedeno ve výsledcích přesnosti. Takto nasbíraná data byla poté manuálně analyzována a bylo z nich zjištěno jaké je procento úspěšnosti predikce skóre, jaké je procento úspěšnosti predikce autora a všechna skóre zdrojů, autorů a zpráv. Při manuální klasifikaci bylo sbíráno vyplnění jednotlivých polí, jako např. titulek neobsahuje clickbait atp. Tato data jsou poté v závěru zmíněna jako potencionální vylepšení detekčního systému do budoucna, kde by mohla být využita právě pro predikci všech příznaků jednotlivých zpráv. Nejpravděpodobněji s využitím vyhodnocených příznaků zdroje a příznaků nejpravděpodobnějšího autora ze stejného zdroje. S využitím autorova průměru (nebo nějakého sestaveného a pozorovaného chování autora) by mohla být sestavena pravděpodobná predikce všech příznaků.

V poslední fázi byla vyhodnocená a zanalyzovaná data ještě ověřena. Tím je myšleno, že byl zkontrolován proces získání těchto dat. Bylo zkontrolováno, zda nedošlo k syntaktické chybě (např. špatně napsaná posloupnost podmínek, nebo špatně zapsané znaky v podmínkách) při zápisu SQL dotazů pro získávání specifických dat.

6 Závěry a doporučení

Z těchto výsledků je patrné, že pro predikci autorství u zpráv je zapotřebí daleko větší množství dat, ze kterých lze lépe rozeznat autory. Také je potřeba mít seznam všech autorů a jejich zpráv. Tudíž při přesnosti predikce autorů je necelých 30% nedostatečné a nelze tedy predikovat autory při 1.200 klasifikovaných zpráv. Hlavním důvodem by mohlo být to, že při tomto množství klasifikovaných zpráv nejsou získáni zdaleka všichni autoři, kteří by mohli být u zprávy uvedení. Při provádění klasifikace, ale lze říct, že při klasifikaci pouze autora ČTK by měl tento model daleko vyšší přesnost, než které aktuálně nabýval, je to předně z toho důvodu, že je autorství pro ČTK uváděno u spousty zpráv.

Predikce skóre zpráv je na základě manuálně ověřených zpráv více než obstojné. Přesnost nad 85% při klasifikovaných 1.200 zpráv (viz 4.3) potvrzuje funkčnost predikce skóre, kde je předpoklad, že by se s dalšími klasifikovanými zprávami přesnost zvyšovala. Je však důležité připomenout, že by bylo vhodné pro zvýšení přesnosti a většího rozsahu dat pro analýzu zahrnout do predikce i autora zprávy, bohužel však spoustu zpravodajských serverů autora v RSS kanálu neuvádí a nelze s nimi tedy pracovat v jednodušší míře. Zároveň by bylo vhodnější zvolit pro strojovou analýzu kompletní zprávy namísto jejich úryvků, ale pro to by se musela zvolit jiná metoda získávání zpráv. Příkladem je web-scapping, který ale nemusí být jednoduše implementovatelný, například kvůli tomu, že má každý zpravodajský server různě postavené webové stránky. A hlavně i z legálních důvodů, kde by byl mnohdy zapotřebí souhlas.

Největší slabinou modelu je s nejvyšší pravděpodobností lidský úsudek, který může při manuální klasifikaci u některých zdrojů ovlivnit příznaky i výsledné hodnocení. Je proto potřeba manuální klasifikaci za pomoci lidských zdrojů maximálně eliminovat. Případnou možností by mohlo být specifikování takových kritérií klasifikace, které by byly natolik specifické, že by v nich nemohl lidský úsudek hrát žádnou roli.

Další částí, která by byla třeba ještě zvážit, a případně přehodnotit, jsou příznaky. Možná by bylo vhodnější zvolit jiné, nebo je ještě rozvést a přidat další. Například by bylo vhodné zahrnout příznak typu komerční zpráva (během manuální klasifikace bylo odhaleno několik takových zpráv). Do určité míry by bylo vhodné nějakým příznakem i technický stav zpravodajského serveru, pokud se jedná o zastaralé neudržované stránky, kde některé odkazy nefungují, určitě by to nepřidalo na důvěře. A takto by se dalo najít určitě spousta dalších příznaků, které by šly zahrnout do hodnocení.

Systém by se mohl dále rozvíjet v predikci jednotlivých příznaků u zprávy na základě již vyhodnocených zpráv. Samozřejmě v první řadě podle kompletních a zprůměrovaných dat zdroje zprávy a autorova hodnocení a průměru jeho příznaků. Díky takové predikci by model mohl dokázat transparentně informovat čtenáře, co s nejvyšší pravděpodobností zpráva obsahuje a na co si dát pozor, nebo případně si dohledat. Pokud by se u této predikce vytvořila vysoká přesnost, systém by měl být schopen dokonce předpovědět příznaky zprávy, kterou autor ještě nepublikoval. Zde by ale bylo třeba sledovat vývoj příznaků u zpráv autora v rámci časového okna, na základě, kterého by mohl být utvořený prediktivní model chování autora. Avšak u některých serverů jsou autoři např. Redakce, kde je predikce chování daleko komplikovanější, protože za tento profil může vydávat i několik autorů najednou.

Tento model by mohl být například využit u nějakého rozšíření do prohlížečů, které by po otevření určitého článku na zpravodajském webu informoval o skóre toho článku a jednotlivých bodech, jak k tomu model došel.

7 Seznam použité literatury

[1] WATZLAWICK, Paul. Jak skutečná je skutečnost?: mylné představy, klamání, porozumění. Hradec Králové: Konfrontace, [cit. 30.3.2023], 1998. ISBN 80-86088-00-6.

[2] TÁBORSKÝ, Jiří. V síti (dez)informací: proč věříme alternativním faktům. Praha: Grada Publishing, [cit. 30.3.2023], 2020. ISBN 978-80-271-2014-7.

[3] Výroční zpráva Bezpečnostní Informační služby za rok 2020 [online]. Bezpečnostní Informační služba, 2021, [cit. 30.3.2023], Dostupné z: https://www.bis.cz/vyrocni-zpravy/vyrocni-zprava-bezpecnostni-informacni-sluzby-za-rok-2020-158d1414.html

[4] GILBERT, Ben. The 10 most-viewed fake-news stories on Facebook in 2019 were just revealed in a new report [online]. [cit. 20.4.2023], 2019. Dostupné z: https://www.businessinsider.com/most-viewed-fake-news-stories-shared-on-facebook-2019-2019-11

[5] ŠLERKA, Josef. Zprávy roku 2021 podle popularity na českém Facebooku [online], 2021, [cit. 30.3.2023]. Dostupné z: https://www.investigace.cz/zpravy-roku-2021/

[6] HAIGH, M., HAIGH, T., & KOZAK, N. I. (2017). Stopping fake news: The work practices of peer-to-peer counter propaganda. Journalism Studies, [cit. 30.3.2023], 1-26. doi:10.1080/1461670X.2017.1316681

[7] ALLCOTT, H., & GENTZKOW, M. (2017). Social Media and Fake News in the 2016 Election. [cit. 30.3.2023]. Cambridge: National Bureau of Economic Research, Inc.

[8] MALÁ, T. Archeolog zjistil, kdy nastane skutečný konec světa podle mayského kalendáře. Je to překvapivé [online]. [cit. 30.3.2023]. Dostupné z: https://www.dotyk.cz/magazin/maysky-kalendar-konec-sveta/

[9] KUDRNA, T. Co dokáže lež. [Dokumentární film]. Praha: Česká televize. 2017, [cit. 30.3.2023]. Dostupné z: http://www.ceskatelevize.cz/porady/11238050887-co-dokaze-lez

[10] BERGHEL, H. Alt-News and Post-Truths in the "Fake News" Era. Computer, (4), 2017, [cit. 30.3.2023], 110. doi:10.1109/MC.2017.104

[11] NOVÁK, O. Proč lidé věří falešným zprávám? Nechtějí vypadnout z kolektivu, tvrdí vědci [online]. 2017, [cit. 30.3.2023]. Český rozhlas Plus. Praha: Český rozhlas. Dostupné z: http://www.rozhlas.cz/plus/dnesniplus/_zprava/proc-lide-veri-falesnym-zpravam-nechtejivypadnout-z-kolektivu-tvrdi-vedci--1714096

[12] CHEN, Y., CONROY, N., RUBIN, V. (2015). News in an Online World: The Need for an "Automatic Crap Detector". Proceedings Of The Association For Information Science & Technology, [cit. 30.3.2023]. 52(1), 1

[13] Argumentační fauly [online]. Bez faulu, [cit. 20.4.2023]. Dostupné z: https://bezfaulu.net/argumentacni-fauly/

8 Seznam obrázků

9 Seznam ukázek kódu

Univerzita Hradec Králové
Fakulta informatiky a managementu

Zadání bakalářské práce

Autor:	**Matěj Habr**
Studium:	I2000356
Studijní program:	B1802 Aplikovaná informatika
Studijní obor:	Aplikovaná informatika
Název bakalářské práce:	**Automatizovaná detekce falešných zpráv**
Název bakalářské práce AJ:	Automated detection of fake news

Cíl, metody, literatura, předpoklady:

Cíl:
Cílem bakalářské práce je seznámení se s problematikou automatizované detekce falešných zpráv (tzv. fake news) spolu s navržením vlastního řešení s využitím strojového učení.

Obsah:

1. Úvod
2. Možnosti detekce fake news, přístup k datům
3. Návrh vlastního systému (princip funkce, struktura)
4. Shrnutí výsledků (použitelnost, účinnost)
5. Závěry a doporučení (další možnosti pro zlepšení detekce)
6. Seznam použité literatury
7. Přílohy

- Detecting Fake News Using NLP Methods, https://dspace.cvut.cz/bitstream/handle/10467/86030/F3-DP-2020-Rehacek-Denis-thesis.pdf?sequence=-1&isAllowed=y

- Thasniya, K. P., et al. Fake News Detection and Prediction Using Machine Learning Algorithms. IJCSIT, Vol. 12(3), 2021, pp. 81-85.

- Mlis, D. G. Apps, AI, and Automated Fake News Detection. Information Outlook V23 N02, 2019. URL: https://www.sla.org/wp-content/uploads/2019/04/IOMarApr2019-FakeNewsDetection.pdf

- WATZLAWICK, Paul. Jak skutečná je skutečnost?: mylné představy, klamání, porozumění. Hradec Králové: Konfrontace, 1998. ISBN 80-86088-00-6.

- NUTIL, Petr. Média, lži a příliš rychlý mozek: průvodce postpravdivým světem. Praha: Grada, 2018. ISBN 978-80-271-0716-2.

- GREGOR, Miloš a Petra VEJVODOVÁ. Nejlepší kniha o fake news, dezinformacích a manipulacích!!!. 2. vydání. Brno: CPress, 2018. ISBN 978-80-264-2249-5.

- MCINTYRE, Lee C. Post-truth. Cambridge, Massachusetts: MIT Press, [2018]. MIT Press essential knowledge series. ISBN 978-0-262-53504-5.

- TÁBORSKÝ, Jiří. V síti (dez)informací: proč věříme alternativním faktům. Praha: Grada Publishing, 2020. ISBN 978-80-271-2014-7.

- Khanam, Z., et al. Fake News Detection Using Machine Learning Approaches. IOP Conf. Ser.: Mater. Sci. Eng. 2021. doi:10.1088/1757-899X/1099/1/012040. URL: https://iopscience.iop.org/article/10.1088/1757-899X/1099/1/012040

- MATTHEW A. RUSSELL. Mining the Social Web. O'Reilly, 2011. ISBN 978-1-449-38834-8.

- Introducing speech and language processing, Coleman, John (John S.), 2005, ISBN 0-521-53069-5.

Zadávající pracoviště:	Katedra informačních technologií, Fakulta informatiky a managementu
Vedoucí práce:	Ing. Martina Husáková, Ph.D.
Datum zadání závěrečné práce:	15.10.2021