

**Česká zemědělská univerzita v Praze**

**Provozně ekonomická fakulta**

**Katedra informačních technologií**



## **Bakalářská práce**

**Porovnání výskytu propagandy na Twitteru a  
Telegramu**

**Jiří Hlavatý**

© 2023 ČZU v Praze

## ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Jiří Hlavatý

Informatika

Název práce

**Porovnání výskytu propagandy na Twitteru a Telegramu**

Název anglicky

**Comparison of the incidence of propaganda on Twitter and Telegram**

---

### Cíle práce

Hlavní cíl:

- Hlavním cílem práce je porovnat četnost výskytu propagandy na sociálních sítích Twitter a Telegram.

Díličí cíle:

- Identifikovat vhodnou metodu k detekci propagandy na sociálních sítích Twitter a Telegram.
- Aplikovat vybranou metodu na datových souborech získaných z Twitteru a Telegramu.
- Vyhodnotit výsledky a interpretovat zjištění.

### Metodika

Metodika řešení teoretické části diplomové práce bude založena na studiu a analýze odborných informačních zdrojů. Na základě znalostí získaných v teoretické části práce bude v praktické části identifikována vhodná metoda k detekci propagandy na sociálních sítích Twitter a Telegram. Dále bude proveden sběr dat ze sociálních sítích Twitter a Telegram pomocí vhodných nástrojů. Na získaných datech bude aplikována metoda k detekci propagandy. Na základě syntézy teoretických poznatků a výsledků praktické části budou formulovány závěry práce.

## Doporučený rozsah práce

50 stran

## Klíčová slova

Propaganda, Sociální sítě, Twitter, Telegram, Rusko-ukrajinská válka, Manipulace, Dezinformace

---

## Doporučené zdroje informací

FTOREK, J. *Manipulace a propaganda na pozadí současné informační války*. Praha: Grada, 2017. ISBN 978-80-271-0605-9.

RYCHLAK, Ronald. *Dezinformace*. BizBooks, 2016.

SANOVICH, Sergey. *Computational propaganda in Russia: the origins of digital misinformation*. 2017.

SHU, Kai, et al. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 2017, 19.1: 22-36.

STANLEY, J. *How Propaganda Works*. [elektronický zdroj] /. Princeton: Princeton University Press, 2015. ISBN 9781400865802.

---

## Předběžný termín obhajoby

2022/23 LS – PEF

## Vedoucí práce

Ing. Miloš Ulman, Ph.D.

## Garantující pracoviště

Katedra informačních technologií

Elektronicky schváleno dne 14. 7. 2022

**doc. Ing. Jiří Vaněk, Ph.D.**

Vedoucí katedry

Elektronicky schváleno dne 27. 10. 2022

**doc. Ing. Tomáš Šubrt, Ph.D.**

Děkan

V Praze dne 25. 02. 2023

### **Čestné prohlášení**

Prohlašuji, že svou bakalářskou práci "Porovnání výskytu propagandy na Twitteru a Telegramu" jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu použitých zdrojů na konci práce. Jako autor uvedené bakalářské práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne \_\_\_\_\_



### **Poděkování**

Rád bych touto cestou poděkoval Ing. Miloši Ulmanovi, Ph.D. za odborné vedení práce a přínosné rady.

# Porovnání výskytu propagandy na Twitteru a Telegramu

## Abstrakt

Tato práce se zabývá propagandou, její definicí a způsoby, jak lze s tímto jevem v moderním světě bojovat. V rámci teoretické části jsou představeny nejnovější zjištění na poli boje proti propagandě a dezinformacím. Důkladněji jsou rozebrány transformační modely, jakožto nejslibněji se vyvíjející technologie, a představeny poslední pokusy jejich využití.

Praktická část se věnuje práci s konkrétním vytrénovaným transformačním modelem BERT. Postupně jsou popsány způsoby sběru datových sad, jejich převedení do vhodného formátu, vytvoření označených dat, průběh práce s transformačním modelem a samotná analýza získaných příspěvků.

V závěru jsou rozebrány naměřené hodnoty a představeny překážky a úskalí na které se během práce s modelem narazilo.

**Klíčová slova:** Propaganda, Sociální sítě, Twitter, Telegram, Rusko-ukrajinská válka, Manipulace, Dezinformace

# Comparison of the incidence of propaganda on Twitter and Telegram

## Abstract

This thesis deals with propaganda, its definition and how this phenomenon can be combated in the modern world. The theoretical part presents the latest findings in the field of fighting propaganda and disinformation. Transformational models, as the most promising emerging technologies, are discussed in depth and recent attempts to use them are presented.

The practical part is devoted to working with a specific trained transformation model, BERT. The methods of collecting datasets, converting them into a suitable format, creating labeled data, the workflow with the transformation model and the actual analysis of the obtained contributions are described step by step.

Finally, the measured values are discussed. Obstacles and pitfalls encountered while working with the model are presented.

**Keywords:** Propaganda, Social networks, Twitter, Telegram, Russia-ukraine war, Manipulation, Disinformation

# Obsah

<b>1. Úvod.....</b>	<b>10</b>
<b>2. Cíl práce a metodika .....</b>	<b>11</b>
<b>3. Teoretická východiska .....</b>	<b>12</b>
1.1 Propaganda na sociálních sítí Twitteru a Telegramu .....	12
1.1.1 Twitter.....	12
1.1.2 Telegram .....	12
1.1.3 Propaganda.....	13
1.1.4 Detekce propagandy .....	15
1.2 Metody zpracování přirozeného jazyka .....	16
1.2.1 Vector Space Model.....	16
1.2.2 Modely s hierarchickou reprezentací .....	19
1.2.3 Transformátory .....	20
1.2.4 BERT .....	22
1.2.5 RoBERTa.....	24
1.3 Přehled přístupů k detekci propagandy .....	24
1.3.1 Hitachi.....	27
1.3.2 ApplicaAI.....	28
1.3.3 aschern .....	29
1.3.4 Ituorp.....	29
1.3.5 ProperGander .....	29
1.3.6 MIC-CIS .....	30
1.4 Sběr dat.....	30
1.4.1 Twitter.....	30
1.4.2 Telegram .....	30
1.5 Shrnutí hlavních poznatků.....	31
1.6 Formulace výzkumné otázky .....	32
<b>4. Vlastní práce.....</b>	<b>33</b>
1.7 Získání datových sad.....	33
1.7.1 Sběr dat z Twitteru.....	33
1.7.2 Výběr omezení při sběru dat z Telegramu .....	34
1.7.3 Sběr dat z Telegramu .....	36
1.8 Začištění dat .....	39
1.8.1 Začištění dat z Twitteru .....	39
1.8.2 Začištění dat z Telegramu.....	39
1.9 Porovnání získaných začištěných datových sad.....	40
1.10 Analýza pomocí transformačního modelu BERT .....	44

1.10.1	Označení dat pro doladění modelu BERT .....	44
1.10.2	Finální doladění modelu BERT .....	45
1.10.3	Vyhodnocení získaných datových sad .....	47
<b>5.</b>	<b>Výsledky a diskuse .....</b>	<b>49</b>
1.11	Výsledky měření .....	49
1.11.1	Získaná data .....	49
1.11.2	Statistická analýza získaných dat .....	51
1.12	Diskuse .....	52
1.12.1	Porovnání výsledků s jinými studii .....	53
1.12.2	Omezení zvoleného přístupu .....	54
1.12.3	Možné rozšíření využití metodiky .....	55
<b>6.</b>	<b>Závěr .....</b>	<b>56</b>
<b>7.</b>	<b>Citovaná literatura .....</b>	<b>58</b>
<b>8.</b>	<b>Seznam obrázků, tabulek, grafů a zkratk .....</b>	<b>63</b>
1.13	Seznam obrázků .....	63
1.14	Seznam tabulek .....	63
1.15	Seznam grafů .....	63
<b>9.</b>	<b>Přílohy .....</b>	<b>64</b>

## 1. Úvod

Jen málokterá síla má v dnešní době moc poštvat proti sobě bratry a sestry, ničit celé státy, svévolně měnit historii a zabíjet nevinné lidi. Přesto existuje úkaz oplývající těmito schopnostmi, propaganda. Ta je v dnešním světě jednou z nejnebezpečnějších věcí, se kterými se moderní společnost musí potýkat. Ačkoli je člověku známá již tisíce let a historicky s ní má nemalé zkušenosti, stále neexistuje spolehlivý způsob, jak se jí bránit, ba ani jak jí identifikovat. Dnes a denně jsme svědky jejích negativních projevů, a to nejen ve světě, ale i v České republice, a častokrát i ve vlastních rodinách.

S prostředky, které propagandě poskytly nové digitální technologie, se její moc ještě navýšila. Našla v nich zbraň schopnou zasáhnout doslova kohokoli kdekoli. Proto je důležité, víc než kdy dříve, studovat jí a najít její podstatu. K tomu, aby tato snaha byla úspěšná je nutné zaměřit se na všechny její aspekty ať už psychologické, sociální, či technické.

Tato práce si neklade za cíl nalézt pověstný Excalibur v boji proti propagandě, ale shrnout nejnovější poznatky v její automatické identifikaci a porovnat její výskyt na konkrétních sociálních sítích Twitter a Telegram na pozadí ukrajinsko-ruské války.

## 2. Cíl práce a metodika

Hlavním cílem této práce je porovnat četnost výskytu propagandy na sociálních sítích Twitter a Telegram. Jednotlivé dílčí cíle byly stanoveny následovně. Nejprve identifikovat vhodnou metodu k detekci propagandy na sociálních sítích Twitter a Telegram. Poté aplikovat vybranou metodu na datových souborech získaných z Twitteru a Telegramu a v konečném kroku vyhodnotit výsledky a interpretovat zjištění. K dosažení těchto záměrů je zapotřebí několika kroků.

V první řadě je pro splnění těchto cílů nutné shromáždit požadované datové sady z Telegramu a Twitteru. Tyto datové sady by se měly co nejvíce shodovat, aby jejich porovnání bylo vypovídající. Z tohoto důvodu bude vybrána jedna konkrétní událost, ukrajinsko-ruská válka, ke které se vybraná data budou vztahovat. Toho bude docíleno vybráním příspěvků spadajících do konkrétního časového úseku a obsahujících vybraná klíčová slova a slovní spojení. Budou brány v potaz jen anglické příspěvky, tak aby se zajistilo, že propaganda cílí na stejné publikum.

Druhým krokem je získání označených dat. Nabízí se dvě možnosti. První je využít některou ze stávajících datových sad. Druhou možností je ruční označení. Ze získaných datových sad z Twitteru a Telegramu budou vybrány náhodné příspěvky a budou rozříděny na propagandistické a ne propagandistické.

Třetím krokem je za pomoci modelu a označených dat provést detekci propagandy v získaných datových sadách.

Konečným krokem bude porovnání dat získaných za pomoci transformačního modelu BERT.

### **3. Teoretická východiska**

#### **1.1 Propaganda na sociálních sítích Twitteru a Telegramu**

Sociální sítě jsou fenoménem dnešní doby. Existují obecné, zaměřené na zveřejňování informací ve formátech všeho druhu. Jako zástupce této skupiny můžeme uvést Facebook a Telegram. Mnoho sociálních sítí se ovšem zaměřuje na specifické oblasti jako je například audiovizuální obsah. Mezi takové sociální sítě patří YouTube, Instagram a v posledních letech populární TikTok. Jako další ukázkou pestrosti, můžeme uvést čistě na obrázky zaměřený Pinterest, nebo ke komunikaci sloužící WhatsApp, Facebook Messenger, Twitter a mnoho dalších. Jak je z tohoto zhrubého přehledu patrné, svět sociálních sítí nabízí širokou škálu přístupů a zaměření.

Je tedy nutné si k pochopení propagandy na Twitteru a Telegramu nejprve definovat čím konkrétně se tyto sítě vyznačují (podkapitola 3.1.1 a 3.1.2) a následně si ve vztahu k nim definovat propagandu a její rysy. (podkapitola 3.1.3 a 3.1.4)

##### **1.1.1 Twitter**

Sociální síť Twitter vznikla počátkem roku 2006 ve firmě Odeo. Základní princip byl poskytnout platformu, kde by uživatelé mohli zveřejňovat krátké příspěvky podobné mobilním zprávám. Nejprve fungovala pouze uvnitř firmy Odeo, posléze, 15. června téhož roku, byla představena široké veřejnosti. V prvních měsících se počet příspěvků, jimž se v rámci sociální sítě Twitter říká „tweety“, pohyboval okolo dvaceti tisíc. O rok později se toto číslo pohybovalo kolem šedesáti tisíc. V roce 2008 na konferenci South by Southwest Interactive se Twitter dostal do širokého podvědomí, a to mu přineslo tři statisíce příspěvků denně. Na počátku roku 2010 to už bylo okolo padesáti miliónů příspěvků. (1)

Podle statistického portálu Statista (2) využívalo v červnu roku 2022 měsíčně Twitter až 436 milionu lidí, což z něj dělá 12. největší sociální síť na světě. Nicméně 9-15 % všech aktivních účtů na Twitteru jsou počítačem řízené účty, takzvaní „boti“. (3) Celkově je v dnešní době denně posláno téměř devět set miliónů příspěvků. (4) Podle serveru datareportal (5) 55.4 % uživatelů Twitteru používá také Telegram.

##### **1.1.2 Telegram**

Sociální síť Telegram vznikla v srpnu roku 2013. Původně se jednalo především o komunikační platformu zaměřenou na komunikaci jeden na jednoho. Uživatelé k Telegramu



přístupují na základě telefonního čísla. Zprávy mezi uživateli jsou uloženy v centralizovaném cloudovém úložišti a lze k nim přistupovat z různých zařízení.

Kromě přímého propojení dvou uživatelů, přidal v září roku 2015 Telegram novou funkci, takzvané vysílací kanály. Každý kanál na Telegramu je nastaven svým vlastníkem tak, aby umožňoval vysílání nebo jednosměrnou komunikaci od malé skupiny odesílatelů k širší skupině uživatelů, příjemců, kanálu. Kromě toho každý kanál může obsahovat přidruženou konverzační místnost, která umožňuje komunikaci mezi všemi účastníky kanálu. Komunikace může být ve formě textu, obrázku, video souboru, audio souboru a podobně.

Výzkum ukázal že 50.1 % příspěvků v kanálech obsahuje přílohu. 53.8 % příloh tvoří fotografie, 29,4 % dokumenty, 16,5 % odkazy na webové stránky a zbylé 0,3 % tvoří hlasování, geografické polohy, hry a kontakty. Dále bylo zjištěno že pouze 2.3 % zpráv obsahuje hashtag, zatímco 27,7 % obsahuje zmínky o jiných uživateli, nebo kanálech. (6)

Podle statistického portálu Statista (2) využívalo v lednu roku 2022 měsíčně Telegram až 550 milionu lidí, což z něj dělá 10. největší sociální síť na světě. Přičemž podle serveru datareportal (5) 61,5 % uživatelů Telegramu používá i Twitter.

### **1.1.3 Propaganda**

V roce 1994 definoval Sproule (7) propagandu jako úsilí organizací, či skupiny, získat na svoji stranu veřejné mínění ve specifické záležitosti, a to za využití masivní propagace atraktivních závěrů, které jsou sestaveny tak, aby zakryly jak svůj přesvědčovací účel, tak nedostatek rozumných důvodů.

Propaganda je v základu rozlišována na „bílou“, „šedou“ a „černou“. „Bílá“ propaganda je založena na informacích odpovídající realitě, kdy je zdroj známý a spolehlivý. „Šedá“ propaganda se vyznačuje nejasným zdrojem informací a tvrzení v ní použítá jsou účelově zkreslena. Častokrát se také opírají o uznávanou autoritu, ačkoli ta prezentovaná tvrzení nikdy neproněsala. „Černá“ propaganda používá zcela smyšlené zdroje a nepravdivá, lživá, tvrzení. Z tohoto důvodu, lze černou propagandu charakterizovat jako dezinformační. (8)

Ve své práci Ellul (9) představil koncept vertikální a horizontální propagandy. Vertikální propaganda vychází především od elit, které při šíření spoléhají na masová média. Velký význam mají komunikační kanály, ve kterých jednotlivec oslovuje masy. Je-li tato

propaganda úspěšná, její dosah se násobí s každým přesvědčeným člověkem, který se automaticky stává dalším propagandistou. Zvláště úspěšný se tento druh propagandy jeví při agitační propagandě, mobilizující davy proti společnému nepříteli. Historickým příkladem může být nacistické Německo. Méně historickým, leč o to aktuálnější příklad můžeme vidět ve spojitosti s ukrajinsko-ruskou válkou, kde je tento druh propagandy umocněn o možnosti poskytnuty digitální komunikací, v čele se sociálními sítěmi.

Protipólem k vertikální propagandě je horizontální propaganda. Ta se opírá o malé autonomní skupiny spolupracující na základě společné ideologie a přesvědčení. Je šířena od spodu společnosti, nikoli z jejího vrcholu. (10)

Historicky byla propaganda vždy spojována se sdělovacími prostředky, jako jsou televize, rádio či noviny. Ačkoli její etymologické kořeny sahají několik století před první sociální sítí, (11) byl to internet, a především sociální sítě, které vnímání propagandy zcela proměnily a neoddělitelně jí spojili s digitálním světem, jakožto místem pro masovou komunikaci.

Nejprve se k šíření propagandy využívali zpravodajské servery, ty však posléze byly nahrazeny sociálními sítěmi, jako jsou právě Twitter a Telegram, které se díky své podstatě stali ideálním místem pro šíření propagandy. Kterýkoli uživatel je na nich schopen šířit pravdivé či nepravdivé informace a manipulovat s nimi tak, aby vyhovovaly jeho vlastním zájmům.

Účinek propagandy na sociálních sítích je dále umocněn samotnými uživateli, kteří ať už chtěně nebo ne, propagandu šíří. Speciálním případem jsou novináři, kteří častokrát sociální sítě využívají k získávání informací a bez ověření je dál předávají prostřednictvím tradičních médií. (12)

Jak ukázal výzkum Troops-Trolls-and-Troublemakers (13) digitální propaganda a její využívání není ojedinělým jevem, ale celosvětovým problémem. Neomezuje se na individuální osoby, ale v široké míře jí využívají i jednotlivé vlády, společnosti a zájmové skupiny. Je založena především na využívání počítačových algoritmů a automatizace k mobilizaci a ovlivňování veřejného mínění. Podle závěrů publikovaných v práci Dissecting a Social Botnet: Growth, Content and Influence in Twitter (14) mohou sociální boti, ať už ve formě takzvané „trollí“ armády, tedy skupiny účtů záměrně šířící dezinformace, nebo jednotlivých anonymních uživatelských účtů, ovlivňovat veřejné mínění mnoha prostředky. Mezi hlavní způsoby ovlivňování veřejného mínění patří astroturfing, (15) což je forma

skryté vertikální propagandy, jejímž cílem je vytvořit dojem horizontální politické podpory, tedy pocitu, že propagovaná myšlenka je ve společnosti všeobecně přijímaná jako správná. Na sociálních sítích zahrnuje kooperaci uživatelských profilů prostřednictvím organizace, například vlády nebo soukromé společnosti, s cílem ovlivnit veřejnost a vzbudit u ní podporu nebo odpor vůči určitému tématu. (10)

Vlády v zemích, jako je Čína a Rusko, v posledních letech ovlivňují politický diskurz na sociálních sítích, právě pomocí astroturfingu. Kromě automaticky šířené propagandy se spoléhají i na placené živé uživatele. (16) (17)

#### **1.1.4 Detekce propagandy**

Detekce propagandy na sociálních sítích je velmi obtížný úkol. Podle J. Farkasa a Ch. Neumayera (10) je základním kamenem k rozpoznání propagandy, ať už skryté či ne, lidský úsudek, neboť jakýkoli algoritmus zatím nedokáže adekvátně analyzovat kulturní souvislosti jednotlivých příspěvků. Dále tuto myšlenku rozvíjí, že hlavními aktéry v boji proti propagandistickým příspěvkům by měli být občanské skupiny složené ze samotných uživatelů sociálních sítí. Tyto skupiny by pak aktivně proti propagandě bojovali nahlašování propagandistických zpráv správcům daných sociálních sítí.

Zároveň ovšem autoři dodávají důvody, proč takovéto řešení nikdy nebude dostačující. Samotná podstata sociálních sítí, která je z velké části založena na anonymitě, dovoluje šířitelům propagandy i po zablokování jejich účtů velmi jednoduše pokračovat v práci. I v případech, kdy se závadný obsah ze sítě stáhne v řádech několika hodin od jeho zveřejnění, může v tu dobu už oslovit tisíce uživatelů a naplnit tak svůj účel. (10)

Dalšími překážkami v úspěšné detekci propagandy pouze lidským úsudkem je neschopnost lidí úspěšně propagandu identifikovat (18), i to že některé manipulativní zdroje informací mohou být studovány a identifikovány až zpětně, poté co se objeví historické dokumenty. (10)

Ačkoli existují indicie naznačující větší přichylnost daného příspěvku k propagandě, nelze na jejich základě vytvořit jasný klasifikační systém. Těmito indiciemi mohou být například délka příspěvku, kdy bylo zjištěno že propagandistické příspěvky na sociálních sítích mají větší délku než ne propagandistické. (19)

Na základě těchto poznatků se detekce propagandy nikdy nebude moci omezit pouze jen na lidský úsudek a je zapotřebí využít některý ze způsobů automatické detekce

propagandy. Tím je myšlena strojová předpověď toho, jestli je daný příspěvek propagandistický, a to bez zásahu člověka, jen za pomoci algoritmů. (18)

## 1.2 Metody zpracování přirozeného jazyka

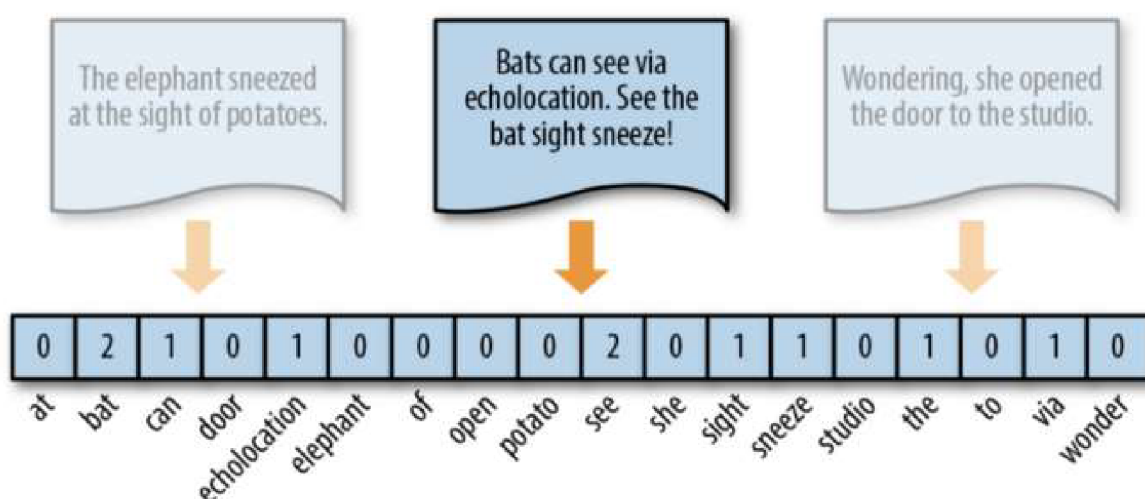
### 1.2.1 Vector Space Model

Jakákoli automatická práce s textem, vyžaduje v první řadě využití některé z technik pro zpracování přirozeného jazyka. V současné době se nejčastěji používají modely reprezentace textu pomocí číselného vektoru. Tyto modely jsou souhrnně označovány jako Vector Space Model, zkráceně VSM.

Převod textu na číselné vektory umožňuje porovnávat ho s jiným textem, a to za pomoci výpočtu různých jader. Jednoduchým a běžně používaným jádrem je normalizovaný vnitřní součin vektorů, který je také známý jako kosinová podobnost. (20)

Jedním z běžně používaných a základních VSM je Bag of Words model, zkráceně BOW. (21) Jeho základem při vytváření jednotlivých vektorů je množina všech slov vyskytující se v daném korpusu dokumentů, přičemž každému slovu je přiřazen unikátní index. Jednotlivé části, povětšinou věty, jsou pak převedeny na vektor tak, že se vytvoří nulový vektor o délce shodné s množstvím všech unikátních slov v korpusu. Pro každé slovo vyskytující se v dané kódované části se na místo, určené podle indexu slova, přičte jedna.

Obrázek 1 - Vizualizace modelu Bag-Of-Words (Zdroj: (22))



Vylepšenou verzí modelu BOW je model Term Frequency - Inversed Document Frequency, zkráceně TF-IDF. Tento model zachycuje rozložení a důležitost jednotlivých

slov v korpusu dokumentů. Tato důležitost je přímo úměrná počtu výskytů slova v dokumentu, ale je kompenzována počtem dokumentů v korpusu, které toto slovo obsahují.

Frekvence termínů (TF): je bodové hodnocení frekvence slova v aktuálním dokumentu. Vzhledem k tomu, že každý dokument je jinak dlouhý, je možné, že se termín objeví vícekrát v dlouhých dokumentech než v kratších. Proto se kvůli normalizaci frekvence termínů dělí délkou dokumentu.

Obrázek 2 - Vzorec pro výpočet TF (Zdroj: (23))

$$TF = \frac{\text{Number of times a word "X" appears in a Document}}{\text{Number of words present in a Document}}$$

Inverzní četnost dokumentu (IDF): je bodové hodnocení toho, jak vzácné je slovo napříč dokumenty. Je-li termín vzácnější, je skóre IDF vyšší.

Obrázek 3 - Vzorec pro výpočet IDF (Zdroj: (23))

$$IDF = \log \left( \frac{\text{Number of Documents present in a Corpus}}{\text{Number of Documents where word "X" has appeared}} \right)$$

Finální skóre je pak vypočítáno jako  $TF * IDF$ . (22)

Jak model BOW tak model TF-IDF má řadu omezení, především ignorují pořadí slov v dokumentu, nerozeznávají synonymní a polysémní slova a nedokáží zachytit sémantický význam jednotlivých částí. Vedle těchto dvou omezení je problémem také to, že s rostoucí velikostí slovníku roste i vektorová reprezentace dokumentů. Výsledkem je vektor se spoustou nulových skóre, tzv. řídký vektor nebo řídká reprezentace, což má za následek větší nároky na paměť a výpočetní zdroje při modelování.

Latentní sémantické indexování, zkráceně LSI, bylo navrženo k eliminaci problémů se synonymními a polysémními slovy. (23) Jedná se o vylepšený BOW model, který umožňuje zachytit sémantický význam pomocí lineární kombinace slov vypočítanou za pomoci singulárního rozkladu hodnot. Opět zde ovšem vzniká problém s přílišnou komplexitou, která omezuje využití LSI. (24) Na základech LSI vznikl model Pravděpodobnostního latentního sémantického indexování, zkráceně PLSI, který vylepšuje

původní model tím, že mapuje dokumenty i slova do reprezentace v tzv. latentním sémantickém prostoru. Toto nové na-mapování dovozuje spolehlivěji odhadnout podobnosti mezi dokumenty nebo mezi dokumenty a slovy než v původní reprezentaci. (25)

LSI ani PLSI ovšem neřeší problémy s rozpoznáváním kontextu a komplexita těchto modelů stále zůstává vysoká. Tyto nedostatky řeší neuronové vkládání slov, v angličtině Neural Word Embedding. Pomocí tohoto řešení dochází ke snížení velikosti za pomoci hustější reprezentace a kontextové podobnosti.

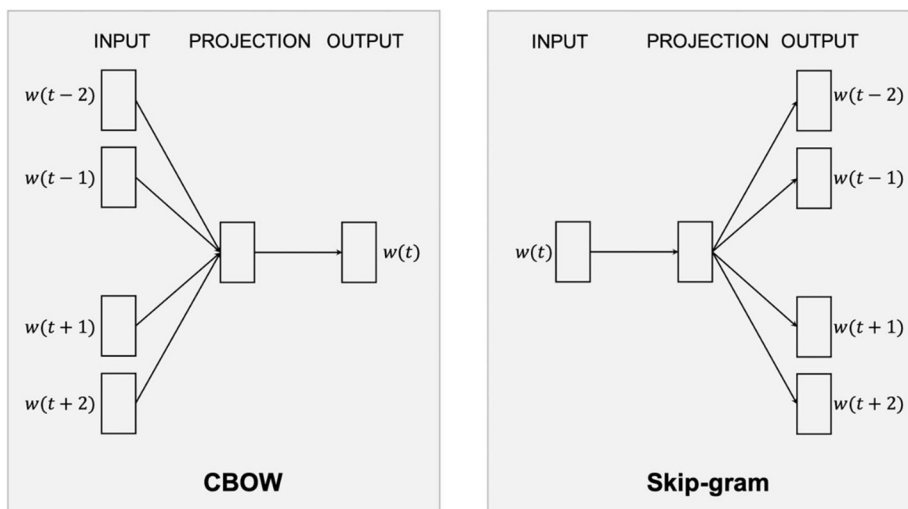
Vkládání slov je reprezentace slov vektory s reálnou hodnotou, přičemž slova, která mají stejný význam, mají podobnou reprezentaci. Příkladem může být rovnice „Paříž – Francie + Česko = Praha“ v tomto případě je rozdíl vektorů slov „Paříž“ a „Francie“ velmi blízký vektoru zachycující význam „hlavní město“. (26)

Základní, myšlenka pro tento přístup je použití hustě distribuované reprezentace pro každé slovo, to je reprezentováno hustým vektorem s reálnou hodnotou o rozměrech v řádech desítek až stovek. Na druhou stranu modely pracující s řídkými vektory, jako jsou BOW a TF-IDF, musejí pracovat s vektory o rozměrech v tisících až miliónech.

Mezi nejpoužívanější modely pro vkládání slov patří Word2Vec a GloVe. (22) Existují dvě architektury modelu Word2Vec. První typ je pokračující „Bag-of-Words“, uváděn pod anglickou zkratkou CBOW. Tento typ Word2Vec modelu předpovídá slova na základě okolních slov. Vstupem jsou v tomto případě vektory okolních slov, respektive jejich průměr. Pořadí jednotlivých slov nemá žádný význam na výsledek a všechny slova mají stejnou váhu. Nejlepších výsledků CBOW dosahuje při použití osmi slov, čtyř předcházejících a čtyř následujících předpovídanému slovu.

Druhý typ je pokračující Skip-gram, ten je podobný CBOW s tím rozdílem, že místo predikce aktuálního slova na základě slov okolních, se snaží předpovědět slova na základě jiného slova ve stejné větě. Přesněji řečeno, každé aktuální slovo je použito jako vstup do log-lineárního klasifikátoru se spojitou projekční vrstvou, který předpovídá slova v určitém rozsahu před a za aktuálním slovem. Bylo zjištěno že čím větší je rozsah tím lepší jsou výsledky, ovšem za cenu zvyšující se komplexity. Jelikož vzdálenější slova jsou povětšinou méně relevantní je jejich váha při výpočtu snížena. (27) (28)

Obrázek 4 - Diagram fungování CBOW a Skip-gram typů modelu Word2Vec (Zdroj: (28))



Model GloVe, zkratka pochází z anglického Global Vectors, je oproti Word2Vec zaměřen na pravděpodobnosti společného výskytu slov v korpusu dokumentů s cílem začlenit je do smysluplných vektorů. Je pro něj důležité, jak často se slovo objevuje v kontextu jiného slova i v rámci celého korpusu. (29)

Neuronové modely pro vkládání slov ovšem stále neřeší problémy s polysémními slovy. Ačkoli jsou vhodné pro zachycení sémantických významů slov, nevyhovují úlohám, ve kterých je potřeba porozumět specifickým literárním formám. Příkladem může být doplnění chybějícího slova ve větě „Počasí bylo krásné, ale foukal \_\_\_\_ vítr“. K předpovědi chybějícího slova je kromě znalosti okolních slov a kontextu nutné taktéž porozumět tomu že spojka "ale" zavádí kontrast mezi větami, takže nejenže chybějící slovo musí odpovídat vhodnému příslovci ke slově „vítr“, ale taktéž musí být v kontrastu ke slově „krásné“.

Zde zmíněné nedostatky jsou důvodem proč modely založené pouze na technologii VSM jsou nevhodné k tak komplexnímu úkolu, jakým je automatická detekce propagandy.

### 1.2.2 Modely s hierarchickou reprezentací

Dalším krokem ve vývoji technik pro zpracování přirozeného jazyka byl přesun od jednovrstvých modelů, tak jak byly výše popsány, k natrénovaným modelům s hierarchickou reprezentací. Tyto vícevrstvé modely se ukázaly velmi efektivními při řešení nejrůznějších úloh související se zpracováním přirozeného jazyka jako je klasifikace textů, zodpovídání otázek, odvozování přirozeného jazyka, řešení referenční shody výrazů a

označování sekvencí. Mezi modely spadající do této kategorie řadíme například ULMFiT, ELMo a na transformátorech založené modely. (22)

Jak ukázaly nejnovější studie a výsledky workshopů, zaměřených na automatickou detekci propagandy, jako byly „NLP4IF-2019 Shared Task on Fine-Grained Propaganda Detection“ a „SemEval 2020 - Task 11: Detection of Propaganda Techniques in News Articles“ nejlepších výsledků při řešení úloh na detekci propagandy v textu dosahovali řešení založené na architektuře transformátorů, jako jsou modely BERT a jeho vylepšená verze RoBERTa. (30) (31)

Model ELMo se od tradičních způsobů vnoření slov liší v tom, že každé slovo reprezentuje funkce celé vstupní věty. Tím, na rozdíl od jiných způsobů vnoření slov jako jsou GloVe a Word2Vec, rozlišuje kontext daného slova. (32)

Metoda Universal Language Model Fine-Tuning, zkráceně ULMFiT, byla představena v práci (33). Stejně jako ELMo je založena na Long short-term memory networks architektuře, zkráceně LSTM. Hlavní výhodou tohoto modelu je jednodušší trénování a menší náročnost na vstupní data, a to bez snížení výkonosti.

### 1.2.3 Transformátory

Transformátory byly poprvé představeny v práci Attention Is All You Need z roku 2017. (34) V dnešní době jsou považovány za nejlepší možná existující řešení pro problémy spojené s automatickým zpracováním textu. Architektura Transformátorů vychází z modelů pro přenosy nervové sekvence. Stejně jako u nich jsou i u Transformačního modelu základem dvě složky, enkodér a dekodér.

Enkodér se skládá ze šesti identických vrstev. Každá vrstva obsahuje dvě podvrstvy. První z nich tvoří vícehlavá sebezpozorovací funkce, v angličtině označována jako „self-attention“. Ta vytváří vektor reprezentující vztah každého slova k ostatním slovům a to za pomoci mechanismu mapování dotazu a množiny dvojic klíč-hodnota na výstup. Všechny klíče, hodnoty a dotazy jsou vektory a pocházejí ze stejného místa, v tomto případě z výstupu předchozí vrstvy enkodéru. Každá pozice v enkodéru může přistupovat na všechny pozice v předchozí vrstvě enkodéru.

Druhou vrstvu enkodéru tvoří jednoduchá, polohově plně propojená dopředná, v angličtině „feed-forward“, síť, skládající se ze dvou lineárních transformací s ReLU

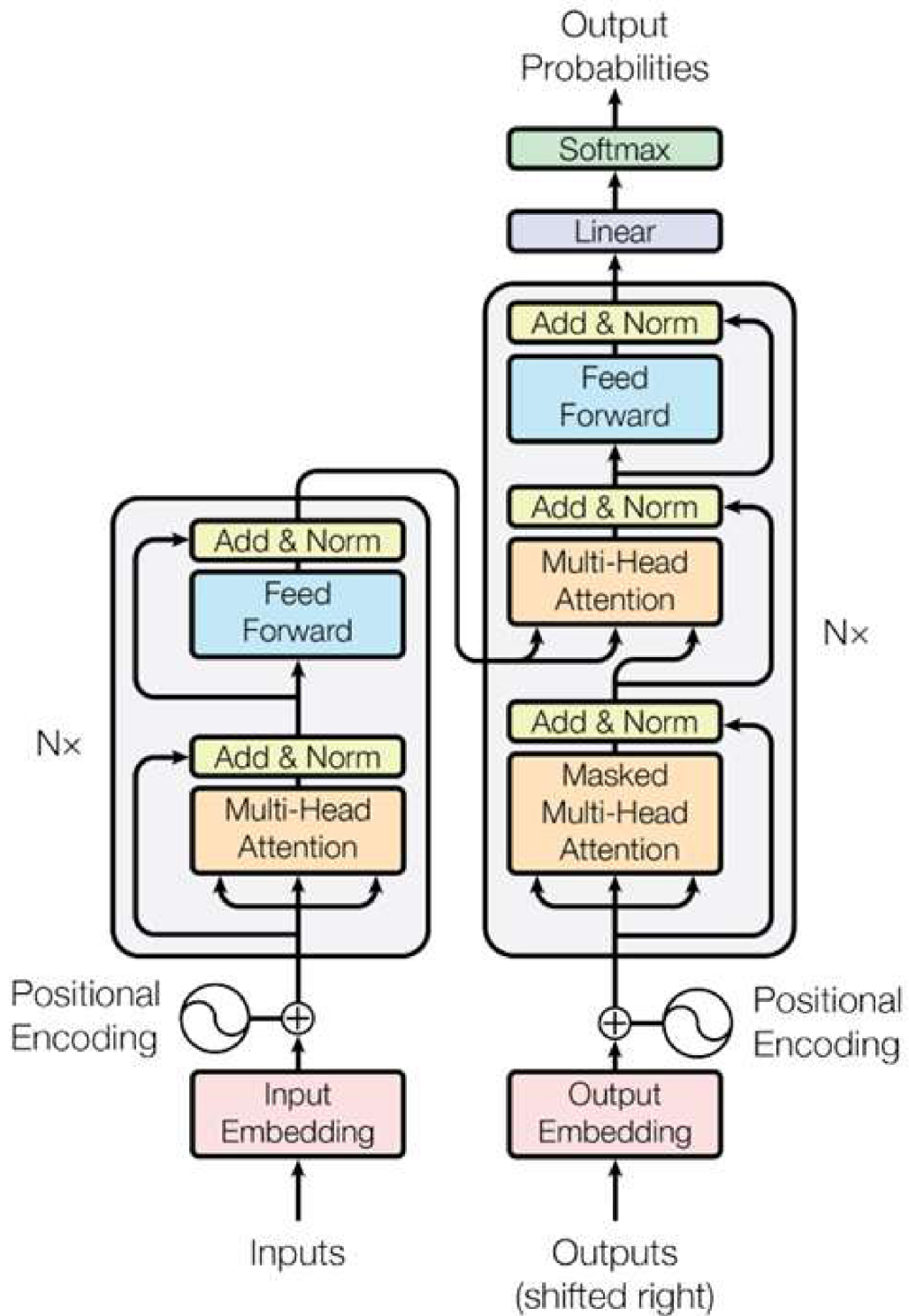


aktivací mezi nimi. Kolem každé podvrstvy je použito zbytkové propojení, které pomáhá síti trénovat tím, že umožňuje přímý tok gradientu. To je následované normalizací vrstev, sloužící ke stabilizaci sítě, což vede k podstatnému zkrácení doby potřebné k natrénování celého modelu. (35) Výstupem každé podvrstvy je  $\text{LayerNorm}(x + \text{Sublayer}(x))$ , kde  $\text{Sublayer}(x)$  je funkce implementovaná samotnou podvrstvou.

Dekodér se taktéž skládá ze šesti identických vrstev. Kromě dvou podvrstev v každé vrstvě enkodéru, dekodér přidává třetí podvrstvu, která provádí vícehlavou pozornostní funkci nad výstupem z enkodéru. Podobně jako u enkodéru je kolem každé podvrstvy použito zbytkové propojení následované normalizací. Sebeopozorovací podvrstva je zde upravena tak, aby se zajistilo že předpovědi budou záviset jen na předchozích známých výstupech. (34)

Na základech Transformátorů byly představeny sebevzdělávající se modely určené k reprezentaci jazyka jako jsou BERT, RoBERTa, XLM a XLNet.

Obrázek 5 - Architektura Transformačních modelů (Zdroj: (35))



#### 1.2.4 BERT

Reprezentační jazykový model BERT, zkratka anglického „Bidirectional Encoder Representations from Transformers“, byl představen v práci „BERT: Pre-training of Deep

Bidirectional Transformers for Language Understanding“ (36). Je navržen tak, aby ho bylo možné natrénovat na porozumění a reprezentaci textu, a to v obou směrech jeho kontextu. Takto vytrénovaný model je poté možno finálně doladit za použití pouze jedné další vrstvy, k řešení široké škály úloh. Mezi ně patří odpovídání na otázky, doplňování chybějících slov anebo detekce specifických výrazů a frází.

Toto zjednodušení a universálnost šetří oproti jiným specificky orientovaným modelům čas, a to bez ztráty výkonosti. Bylo dokázáno že BERT dosahuje stejného výkonu jako nejmodernější specificky zaměřené modely, a to v široké škále úloh. (36)

Učení modelu BERT se skládá ze dvou po sobě jdoucích fází. První je fáze natrénování, označovaná také jako „pre-training“, a po ní následuje fáze doladění, označovaná jako „fine-tuning“.

Během natrénování je model učen na neoznačených datech za použití dvou trénovacích úloh. První úloha využívá maskovaný jazykový model, zkráceně MLM z anglického „Masked Language Model“, který je inspirován úlohou Cloze (37). MLM náhodně maskuje některé ze slov s cílem předpovědět tyto maskovaná slova pouze na základě jejich kontextu a to v obou směrech. Právě zahrnutí jak pravolevého, tak levoprávého kontextu slova je jednou z hlavních výhod BERT modelu. Druhá úloha je založena na predikci následující věty. Kdy je z korpusu vybrána věta A a věta B. V polovině případů je věta B větou, která následovala větu A a je tak označena jako následující, v druhé polovině se jedná o náhodnou větu z korpusu a je označena jako ne-následující. Pro natrénování byly využity korpusy BooksCorpus, obsahující 800 milionů slov a anglická Wikipedia, obsahující 2,5 miliardy slov.

Při doladění je BERT model nejprve spuštěn s natrénovanými parametry a pomocí označených dat jsou tyto parametry doladěny. Tento postup umožňuje vzít jeden natrénovaný BERT model a využít ho pro různé úlohy jen doladěním parametrů poskytnutím označených dat. Aby BERT mohl zvládnout rozdílné úlohy, mohou být jako vstupy k učení použity jednotlivé segmenty, nejčastěji věty, nebo dvojice takovýchto segmentů.

BERT model byl v základu natrénován ve dvou velikostech BERT<sub>BASE</sub>, využit hlavně k porovnání se stejně velkým OpenAI GPT modelem, a BERT<sub>LARGE</sub>. Jednotlivé varianty se liší v nastavení hodnot modelu. Konkrétně se jedná o počet vrstev (L), velikost skryté vrstvy (H), počet hlav u sebezpozorovací vrstvy (A) a celkový počet parametrů (P).

BERT<sub>BASE</sub> byl nastaven v konfiguraci L=12, H=768, A=12, P=110 milionů. BERT<sub>LARGE</sub> v konfiguraci L=24, H=1024, A=16, P=340M (36)

### 1.2.5 RoBERTa

Model RoBERTa, zkratka vychází z anglického „Robustly optimized BERT approach“ a vznikl v návaznosti na model BERT. Hlavní zaměření při vývoji RoBERTa modelu bylo vyhodnocení vlivu parametrů a velikosti tréninkové sady při ladění modelu. Bylo zjištěno že model BERT není dostatečně natrénován a RoBERTa model tak přichází s několika vylepšeními.

V úloze využívající maskovaný jazykový model byla využita dynamická změna maskovacího vzoru aplikovaného na tréninková data, tím se předešlo využívání stejné masky napříč tréninkovými epochami. Výsledky ukazují že dynamické maskování vykazuje srovnatelné, nebo mírně lepší výsledky než statické maskování.

Dalšími změnami oproti původnímu BERT modelu je odstranění úlohy na predikci následující věty, trénování modelu s více daty a trénování na delších sekvencích. (38)

## 1.3 Přehled přístupů k detekci propagandy

Výzkum v oblasti odhalování propagandy na základě analýzy textu má krátkou historii, a to zejména kvůli nedostatku vhodných označených datových souborů, které jsou nutné pro správný trénink modelů zpracovávající přirozený jazyk.

V poslední době se objevili iniciativy, v jejichž rámci se analyzují zprávy, které následně mohou být použity k tréninku modelů. Například nezávislá organizace Media Bias/Fact Check, která analyzuje média z hlediska jejich faktického zpravodajství, neobjektivitu a propagandistického obsahu. Podobnou iniciativu provozuje i US News & World Report a Evropská unie. (39)

Kromě těchto iniciativ vznikají i datové sady Označené přímo pro trénink transformačních modelů. V práci “A Survey on Computational Propaganda Detection“ (39) jsou uváděny tři, které autoři považují za vhodné.

Prvním je datová sada TSHP-17 (40). Tato sada obsahuje novinové články a řadí je do jedné ze čtyř kategorií: důvěryhodné, satirické, hoax a propagandistické. Označení jednotlivých článků bylo učiněno pomocí vzdáleného dohledu, což znamená že článek je rozřazen do jednotlivých tříd podle toho do jaké třídy ho zařadila některá třetí strana, v tomto

případě se jednalo o zařazení od US News & Report. Jednotlivé články byly shromážděny z „English Gigaword“, jež archivuje novinové články, a ze sedmi nespolehlivých zdrojů.

Druhou zmiňují datovou sadu QProp (41). Autoři poznamenávají že nevýhodou datové sady TSHP-17 je malé množství zdrojů. Sada QProp je proto dvakrát větší a rozděluje články binárně na propagandistické a věrohodné, přičemž přibližně 10 % článků je označeno jako propagandistické. QProp byl znovu označen pomocí vzdáleného dohledu, v tomto případě se jednalo o označení za pomoci serveru Media Bias/Fact Check.

Jelikož jak TSHP-17 tak QProp byly označené pomocí vzdáleného dohledu a označují vždy celé články, není v těchto sadách možné identifikovat přesný propagandistický úryvek. Proto jako třetí sadu uvádí PTC (42). Ta se vyznačuje třemi důležitými odlišnostmi. Zaprvé označení probíhá na úrovni fragmentů, nikoli celých článků. Zadruhé se k označení používají profesionálové a nikoli vzdálený dohled. Za třetí se místo binárního kategorizování na propagandistický a nepropagandistický, úryvky kategorizují na základě 18 různých propagačních technik. Jednotlivé techniky jsou zaznamenány v tabulce „Druhy propagandistické techniky“. (39) Ačkoli rozsahem je PTC menší než TSHP-17 a QProp, obsahuje přes 7000 propagandistických úryvků.

Tabulka 1 - Druhy propagandistické techniky (Zdroj: (43), vlastní překlad)

Technika	Definice
Nadávky	Útočení na cíl propagandy urážlivým označením.
Opakování	Opakování stejné zprávy stále dokola.
Slogany	Používání krátkých a zapamatovatelných frází.
Odvolání se na strach	Podporování myšlenek a názorů zasévajících strach z jiných možností.
Pochyby	Zpochybňování důvěryhodnosti někoho nebo něčeho.
Přehánění	Zvětšování nebo zmenšování něčeho.
„Flag-Waving“	Působení na vlastenectví a identitu.
Citově zabarvený jazyk	Působení na emoce a stereotypy.
Zjednodušení „Ad Hitlerum“	Odmítat myšlenku tím, že je připisována skupině, kterou publikum nenávidí.
Stádový efekt	Působení na popularitu myšlenky nebo názoru.
Zjednodušování	Přirazování jednoduchých příčin složitým událostem.

Zamlžování	Používání záměrně nejasného a obskurního vyjadřování, za účelem zmatení publika.
Naléhání na autority	Používání podpory od autority jakožto důkazu.
Černobílý omyl	Prezentování pouze dvou možností namísto více.
Klišé znemožňující myšlení	Používání frází, které odrazují od kritického myšlení a smysluplné diskuse.
Falešná stopa	Prezentování irelevantních materiálů, které odvádějí pozornost.
Strašák	Vyvracení argumentů, které nebyly předneseny.
Whataboutismus	Obviňování oponenta z pokrytectví.

V rámci „NLP4IF-2019 Shared Task on Fine-Grained Propaganda Detection“ se na základě datového souboru PTC vytvářeli modely pro určení druhu a automatickou detekci propagand. Nejúspěšnější modely používaly kontextové reprezentace založené na modelu BERT. Další využívaly RoBERTa, GloVe a ELMo. (31)

V roce 2020 proběhl workshop SemEval, který ve své úloze „Task 11: Detection of Propaganda Techniques in News Articles“ využil stejné novinové články a v rámci dvou pod úloh se zajímal o detekci části textu obsahující propagandu a o kvalifikaci konkrétního typu propagandy. Z důvodů nedostatečného vzorku dat zredukovali 18 kategorií na 14. (30) Stejně jako u NLP4IF-2019 i na SemEval 2020 se nejúspěšnějšími řešeními projevíly ty, které byly založené na modelech BERT a RoBERTa. Dále se zde využívaly XLM, XLM RoBERTa, ALBERT, GPT-2 SoanBERT a DistilBERT modely. Tabulka „Přehled použitých technik na SemEval 2020“ ukazuje přehled jednotlivých řešení. Bílé zaškrtnutí na černém pozadí znamená použitou technologii, černé zaškrtnutí znamená uvažovanou, avšak nepoužitou technologii.

Tabulka 2 – Přehled použitých technik na SemEval 2020 (Zdroj: (31))

Rank. Team	Transformers										Learning Models						Representations						Misc								
	BERT	RoBERTa	XLNet	XLNet	XLNet RoBERTa	ALBERT	GPT-2	SpanBERT	LaserTagger	LSTM	CNN	SVM	Naive Bayes	Boosting	Log regressor	Random forest	CRF	Embeddings	ELMo	NEs	Words/n-grams	Chars/n-grams	PoS	Trees	Sentiment	Subjectivity	Rhetorics	LIWC	Ensemble	Data augmentation	Post-processing
1. Hitachi	☑	☑	☑	☑	☑	☑				☑							☑	☑	☑										☑		
2. ApplicaAI	☑	☑	☑	☑	☑	☑											☑	☑	☑										☑	☑	
3. aschern	☑	☑	☑	☑	☑	☑											☑	☑	☑										☑	☑	☑
4. LTIatCMU	☑	☑	☑	☑	☑	☑				☑	☑						☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
5. UPB	☑	☑	☑	☑	☑	☑											☑	☑	☑										☑	☑	
7. NoPropaganda	☑	☑	☑	☑	☑	☑		☑	☑	☑	☑						☑	☑	☑												
8. CyberWalle	☑	☑	☑	☑	☑	☑				☑	☑		☑				☑	☑	☑						☑	☑	☑	☑	☑	☑	☑
9. Transformers	☑	☑	☑	☑	☑	☑				☑	☑						☑	☑	☑										☑	☑	
11. YNUtaoxin	☑	☑	☑	☑	☑	☑											☑	☑	☑												
13. newsSweeper	☑	☑	☑	☑	☑	☑												☑	☑	☑				☑							
14. PsuedoProp	☑	☑	☑	☑	☑	☑												☑	☑	☑									☑	☑	
16. YNUHPCC	☑	☑	☑	☑	☑	☑				☑	☑						☑	☑	☑										☑	☑	
17. NLFIT	☑	☑	☑	☑	☑	☑				☑	☑						☑	☑	☑										☑	☑	
20. TTUI	☑	☑	☑	☑	☑	☑												☑	☑	☑								☑	☑	☑	☑
21. BPGC	☑	☑	☑	☑	☑	☑												☑	☑	☑											
22. DoNotDistribute	☑	☑	☑	☑	☑	☑				☑	☑							☑	☑	☑									☑	☑	
23. UTMN	☑	☑	☑	☑	☑	☑								☑				☑	☑	☑					☑						
25. syrapropa	☑	☑	☑	☑	☑	☑		☑	☑									☑	☑	☑											
26. SkoltechNLP	☑	☑	☑	☑	☑	☑				☑	☑						☑	☑	☑												
27. NTUAILS	☑	☑	☑	☑	☑	☑				☑	☑							☑	☑	☑											
28. UAIC1860	☑	☑	☑	☑	☑	☑												☑	☑	☑											
31. 3218IR	☑	☑	☑	☑	☑	☑				☑	☑							☑	☑	☑											
33. WMD	☑	☑	☑	☑	☑	☑												☑	☑	☑								☑	☑	☑	☑

V následujících podkapitolách budou rozebrány ty přístupy, které v rámci jednotlivých workshopů dosahovali nejlepších výsledků. U SemEval byl brán v potaz výsledek pod-úkolů zaměřující se na detekci propagandy, jelikož pod-úkol identifikující konkrétní druh propagandy není pro tuto práci relevantní. Podobně u NLP4IF-2019 byly vybrány ty řešení, které dosahovali nejlepších výsledků v úkolu zaměřeném na rozpoznávání propagandy.

### 1.3.1 Hitachi

Tým Hitachi dosáhl v rámci SemEval 2020 nejlepšího výsledku. Své řešení postavili na transformátorech, a to ze dvou hlavních důvodů. Jednak se inspirovali studiemi „Fine-Grained Propaganda Detection with Fine-Tuned BERT“ (43) a „Sentence-Level Propaganda Detection in News Articles with Transfer Learning and BERT-BiLSTM-Capsule Model“ (44), jenž k řešení navrhuje použít vylepšený natrénovaný jazykový model BERT. Zadruhé z důvodu, že detekce propagandy je velice závislá na sémantickém porozumění textu, ke kterému jsou transformátory vhodné.

Kvůli nedostatku studií na toto téma pro nejnovější transformační modely, jako jsou RoBERTa a XLNet, se při testování rozhodli využít celkem šest různých transformačních modelů. Konkrétně se jedná o BERT, GPT-2, RoBERTa, XLM-RoBERTa, XLNet a XLM. Konkrétní natrénované verze jsou uvedeny v tabulce „Přehled verzí využitých transformátorů týmem Hitachi“.

Tabulka 3 - Přehled verzí využitých transformátorů týmem Hitachi (Zdroj: (46))

Transformátor	Využitá natrénovaná verze
BERT	„bert-large-cased-whole-word-masking“
GPT-2	zero-shot setting
RoBERTa	„roberta-large“
XLM-RoBERTa	„xlm-roberta-large“
XLNet	„xlnet-large-cased“
XLM	„xlm-mlm-en-2048“

Empiricky bylo zjištěno následující. Modely RoBERTa a XLNet vykazovali nejlepší výsledky při detekci propagandy. Vytvoření zastřešujícího modelu, který sdružuje všech šest zmíněných transformátorů, dosahuje lepších výsledků, než jsou-li transformátory použity separátně. Každý model má jinou rychlost učení a nalezení optimální rychlosti je zásadní pro získání vysokého výkonu. (45)

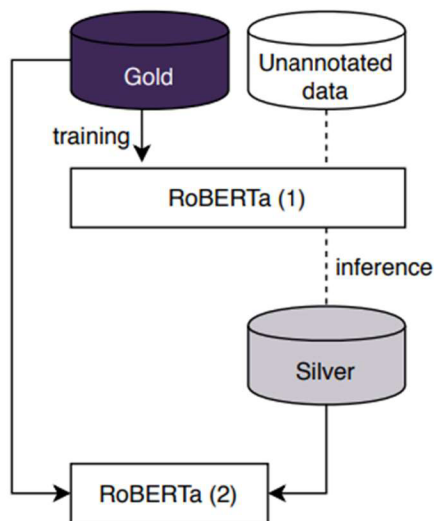
### 1.3.2 ApplicaAI

Tým ApplicaAI se umístil na druhém místě během SemEval 2020. Jejich řešení je postaveno na RoBERTa-CRF modelu, konkrétně na specificky upraveném RoBERTa<sub>LARGE</sub>. Nejprve se tento model natrénoval na poskytnutých označených datech, dokud se nedosáhlo nejlepšího výsledku. Poté se takto natrénovaný model použil k označení půlmiliónu náhodných vět ze serveru OpenWebText. Vzniklá, automaticky označená data, byla spolu s původními daty využita k opětovnému natrénování RoBERTa modelu. Tento postup se opakoval celkem třikrát.

Výsledky dosažené v jednotlivých kolech byly o 0,13%, 1,06% a 2,61% lepší, než výsledky dosažené modelem natrénovaným pouze s původně poskytnutými daty. (46)



Obrázek 6 - Diagram postupu využitý týmem ApplicaAI (Zdroj: (47))



(46)

### 1.3.3 aschern

Tým aschern skončil v rámci SemEval 2020 na třetím místě. K řešení využil dva spojené RoBERTa-CRF modely, každý natrénovaný na jiných datech. Výsledky byly dodatečně zpracovány. Zaprvé proběhla kontrola, že počáteční a konečný znak detekovaného propagandistického úryvku je číslo nebo písmeno. V případě že nebylo, byl úryvek zkrácen. Druhým krokem byla kontrola, jestli úryvku předchází nebo následuje otazník. Pokud tomu tak bylo, byl i on zahrnut do úryvku. (47)

### 1.3.4 Ituorp

Tým ituorp dosáhl NLP4IF-2019 nejlepšího výsledku na testovacích datech. Nejprve pracovali s modelem TF-IDF, ale nakonec nejlepších výsledků dosáhli s BERT modelem za použití dvanácti pozorovacích hlav a dvanácti transformačních bloků. (48)

### 1.3.5 ProperGander

Tým ProperGander, druhý v NLP4IF-2019, využil BERT<sub>BASE</sub> model. Při svém výzkumu zjistili že rozšíření nevyvážených tréninkových dat pomocí převzorkování přineslo zlepšení výkonosti o 4 %. (49)

### 1.3.6 MIC-CIS

Tým MIC-CIS obsadil třetí místo na NLP4IF-2019. Použil sdružení tří přístupů. Logické regrese, konvoluční neuronové sítě a doladěný BERT model. Ve všech případech se používalo vkládání slov za pomoci FastText a natrénovaný BERT model. Přičemž rozhodnutí zdali je segment propagandistický se určilo pomocí hlasování. (50)

## 1.4 Sběr dat

Důležitou součástí této práce je sběr vzorku dat z Telegramu a Twitteru podle co nejpodobnějšího klíče. Následující podkapitoly se zabývají jaké možnosti sběru dat jednotlivé platformy nabízejí.

### 1.4.1 Twitter

Twitter poskytuje velice jednoduchý a přímočarý způsob k automatickému sběru příspěvků. K tomuto účelu existuje speciální API rozhraní. Pro práci s tímto API je nutné nejprve získat ověřený API klíč a autorizační token. Toho lze dosáhnout žádostí o vytvoření vývojářského účtu. Jakmile je požadavek ze strany Twitteru schválen, stačí založit projekt a přiřadit k němu vzorovou aplikaci. Tato aplikace následně vygeneruje jak API klíč, tak autorizační token.

Nejjednodušším způsobem, jak pracovat s Twitterovým API po získání přístupových údajů je pomocí knihovny Tweepy a skriptu napsaném v Pythonu. Nejprve je potřeba si knihovnu nainstalovat a následně pomocí autorizačního tokenu inicializovat klienta, který komunikuje s Twitter API. Dalším krokem je vytvoření vyhledávacího dotazu (51), v němž lze specifikovat jaký výraz, hashtag, frázi, emotikon a mnoho dalšího co má příspěvek obsahovat, celkem je možno nadefinovat třicet šest různých operátorů. Jednotlivé operátory mezi sebou lze kombinovat pomocí logických operátorů. (52)

### 1.4.2 Telegram

Sběrem dat z Telegramu se zabývala studie „The Pushshift Telegram Dataset“ (6). Ačkoli Telegram nenabízí tak otevřené a jednoduché cesty pro sběr dat, byl v této práci identifikován možný postup k získání dostatečného vzorku dat. Autorům se podařilo vytvořit korpus o velikosti 317,224,715 zpráv, získaných z 27,801 kanálů od 2,200,040 unikátních uživatelů. Skript pomocí kterého tento korpus vytvořili zveřejnili a je volně přístupný. (53)

K samotnému sběru dat byla použita Python 3.0 knihovna Telethon, umožňující komunikaci s API Telegramu. (54) Na začátku sběru dat bylo ručně vybráno 261 anglických kanálů z nichž 124 bylo zaměřeno na pravicový extremismus a 137 kanálů bylo spojeno s kryptoměnami. K rozšíření kanálů na finální počet byla využita funkce Telegramu „forwarding“, která dovoluje přeposílat obsah jednoho kanálu na kanál jiný.

Obdobně jako Twitterové API i Telegramové API vyžaduje zřízení vývojářského účtu a získání autorizačního tokenu.

## 1.5 Shrnutí hlavních poznatků

Ač je propaganda světu známá již staletí, ukázalo se, že moderní telekomunikační technologie, v čele s internetem, vytváří ideální podhoubí pro její šíření a využívání. Ze své složité podstaty a z lidské přirozenosti je pro člověka velmi obtížné, ne-li nemožné propagandu automaticky rozpoznat. Dokáže lidi spolehlivě přesvědčit o cílených názorech, proměnit jejich myšlení a chování tak, aby odpovídalo chtěnému způsobu, a to lépe než jakákoli jiná metoda na světě. Tohoto dnes a denně masově využívají různé zájmové skupiny. V první řadě se jedná o státy, přičemž ve světě nenajdeme jedinou zemi, která by ve větší či menší míře nevyužívala některou z technik propagandy k ovlivnění lidí. Kromě států propagandu využívají taktéž různé spolky, firmy, náboženské organizace a další skupiny lidí snažící se prosadit své názory a zájmy.

Jediným možným způsobem, jak proti fenoménu propagandy bojovat je její automatická detekce. Přes veškerou snahu a velké posuny, které v posledních rocích udělala, je automatická detekce propagandy stále ve svých počátcích. Prozatím neexistuje volně přístupný a spolehlivý nástroj pro širokou veřejnost na její rozpoznání, avšak poslední pokusy v tomto úsilí, jako byly akce „NLP4IF-2019 Shared Task on Fine-Grained Propaganda Detection“ a „SemEval 2020 - Task 11: Detection of Propaganda Techniques in News Articles“, přinesli nadějně výsledky.

Jako nejideálnější přístup se ukázaly metody založené na transformačních modelech. Tyto modely, naučené na rozpoznávání textu, které se doladěním na specificky označených datech dají modifikovat pro konkrétní problém, se v posledních letech ukázaly jako nejspolehlivější pro detekci propagandy a můžeme předpokládat, že jejich zdokonalováním a vylepšováním se bude následující vývoj ubírat. Toto řešení taktéž přislubuje, že by se společnost v blízké budoucnosti mohla dočkat spolehlivého, volně přístupného nástroje na

detekci propagandy, založeném právě na natrénovaném, doladěném transformačním modelu.

## **1.6 Formulace výzkumné otázky**

Lze pomocí transformačního modelu BERT určit do jaké míry se liší četnost anglicky psaných propagandistických příspěvků v souvislosti s Ukrajinsko-ruskou válkou, na sociálních sítích Twitter a Telegram?

## 4. Vlastní práce

Vlastní práce je členěna na jednotlivé podkapitoly podle způsobu, jakým je nutno pracovat s transformačními modely. Nejprve se kapitola 4.1. věnuje sbírání datových sad z Telegramu a Twitteru. V kapitole 4.2 je popsán způsob jakým byla data začištěna, aby je bylo možno použít v transformačním modelu. Kapitola 4.3 získaná začištěná data porovnává. Samotné analýze se věnuje kapitola 4.4 a její dílčí podkapitoly.

### 1.7 Získání datových sad

V této kapitole bude rozebrán postup při získávání vhodných datových sad. V podkapitole 4.1.1 bude popsán způsob filtrace vhodných příspěvků a jejich následné sesbírání ze sociální sítě Twitter. Druhá část této kapitoly se věnuje získávání vhodných dat z Telegramu, které se ukázalo náročnější. Z tohoto důvodu je výběru omezení při získávání dat z Telegramu věnována samostatná podkapitola 4.1.2. Sběru samotných dat se věnuje následující podkapitola 4.1.3.

#### 1.7.1 Sběr dat z Twitteru

Jak bylo zmíněno v teoretické části, Twitter dovoluje sbírat data od uživatelů pomocí vlastního API rozhraní. Ačkoli se doporučená knihovna Tweepy jeví jako vhodný způsob, jak data získat, bylo při práci na scriptu, který by využíval tuto knihovnu zjištěno, že finální script nedosahuje požadovaných výsledků. Proto byla nakonec pro sběr dat využita volně dostupná knihovna Snsrape (55). Tato knihovna umožňuje sběr dat z Twitteru bez nutnosti vytvářet si vlastní přihlašovací údaje k nativnímu API rozhraní a veškerá komunikace s tímto API je implementována uvnitř ní. Uživatelovy stačí pouze definovat okruh příspěvků, které se mají shromáždit.

Filtrace jednotlivých příspěvků byla prováděna za pomoci stejných příkazů, jaké jsou využity na platformě Twitter. (56) Konkrétně byly příspěvky omezeny klíčovým slovem, jazykem a datem publikace. Vybrané zprávy musely obsahovat jedno z následujících slov nebo frází: „Russia“, „Ukraine“, „Ukraine War“, „Russian War“, „Putin“ nebo „Zelensky“, datum publikace muselo být 05. prosince 2022 a příspěvek musel být psán anglickým jazykem.

Obrázek 7 - Script pro sběr dat z Twitteru (Zdroj: vlastní)

```
# Import knihoven
import snsrape.modules.twitter as sc
import pandas as pd

# Vytvoření struktur pro uchování získaných dat
df = pd.DataFrame()

# Sběr příspěvků podle filtru
filtr = "(Russia OR Ukraine OR Ukraine War OR Russian War OR Putin OR Zelensky) until:2022-12-05 since:2022-12-04 lang:en"
for t in sc.TwitterSearchScraper(filtr).get_items():

    # Získání všech hashtagů
    h = ""
    if t.hashtags != None:
        for ele in t.hashtags:
            h += ele
            h += ";"

    # Získání informací o datu, hashtagu, autorovi a textu příspěvku
    data = {"sender": t.user.username, "text": t.content, "date": t.date, "hashtags": h}

    # Uložení dat
    temp_df = pd.DataFrame(data, index=[1])
    df = pd.concat([df, temp_df])

# Export dat do tabulky
df['date'] = df['date'].dt.tz_localize(None)
df.to_excel("C:\\Users\\jirih\\Desktop\\Datasey\\Twitter\\TwitterData.xlsx", index=False)
```

Výsledkem tohoto scriptu bylo 15600 shromážděných příspěvků.

## 1.7.2 Výběr omezení při sběru dat z Telegramu

Podle předpokladů byl sběr vhodných dat z Telegramu obtížnější a náročnější. Nejprve byl určen klíč pro výběr dat, na jehož základě byl posléze napsán script v Pythonu.

Z důvodů omezení API Telegramu, která neumožňuje sběr dat napříč všemi kanály, bylo nutno jako první rozhodnout které skupiny či kanály budou vybrány pro shromažďování dat. Hlavním rozdílem mezi skupinou a kanálem je velikost uživatelské základny. Zatímco velikost největších skupin se pohybuje v řádek statisíců, největší kanály jsou až 30x větší. (57)

Pro účely této práce, jenž má za cíl porovnat propagandu na Telegramu, u kterého je dosah příspěvku vždy omezen skupinou či kanálem, s Twitterem, jenž dovoluje téměř neomezený dosah, byly jako základní jednotky pro sběr příspěvků vybrány kanály. Hlavním důvodem byl větší dosah a tím pádem srovnatelnější podmínky mezi oběma sociálními sítěmi.

Druhým krokem bylo vybrání konkrétních kanálů. Byly stanoveny základní podmínky, které musel daný kanál splňovat:

1. Musel být v anglickém jazyce.

2. Muselo se jednat o zpravodajský, informační či novinkový kanál.
3. Musel být alespoň z části zaměřen na Rusko-Ukrajinskou válku.
4. Vybrané kanály by měli být z různého spektra.

**Bod – Kanály v anglickém jazyce:** Byl opět určen na základě účelu, se kterým byla data sbírána, na porovnání s Twitterem. Jelikož na Twitteru byla data sesbírána pouze v anglickém jazyce, a tedy se zaměřením na anglicky mluvící uživatele, vybráním anglicky psaných kanálů se tak docílilo toho, že cílová skupina bude na obou platformách podobná. Další výhodou sesbírání dat pouze v angličtině je, že se ve finální části dají data porovnávat pomocí stejně dotrénovaného modelu, bez nutnosti deformovat je překladem. Nevýhodou tohoto kritéria ovšem je, že většina největších zpravodajských kanálů z oblasti Ruska a Ukrajiny je na telegramu psaná v ruštině (popřípadě ukrajinštině) a diskvalifikuje se tak značné množství jinak vhodných kanálů. I přes tuto nevýhodu bylo pro účely této práce kritérium anglického jazyka zachováno.

**Bod – Kanály zpravodajské, informační či novinkové:** Toto kritérium bylo určeno na základě definice propagandy a toho čeho se snaží dosáhnout.

**Bod – Kanály alespoň z části zaměřené na rusko-ukrajinskou válku:** Kritérium vyplývající ze zadané výzkumné otázky. Tento filtr doplňuje klíčová slova, použita při filtraci Twitterových dat a dovolí porovnat získaná data s daty z Twitteru.

**Bod – Kanály z různého spektra:** Jelikož kanály jsou komunikační prostředky formy jeden ku všem, tak jsou názory na nich povětšinou zaměřeny velice konkrétně se sympatiemi k jedné či druhé straně konfliktu. Ve velké míře je podpora té či oné strany či myšlenky hlavním a jediným důvodem existence jednotlivých kanálů. Tak aby tato práce obsáhla co nejširší spektrum dat, bylo rozhodnuto o vybrání kanálů z obou stran konfliktu. Tedy ty, které vykazující znaky podpory ruské nebo ukrajinské straně a ty které vykazují znaky nezávislého zpravodajství.

Pro vybrání konkrétních kanálů byl využit server TGStat (57) obsahující volně přístupné statistiky o Telegramových kanálech a skupinách. Pomocí funkce „Search by posts“ byl na základě klíče „Special Military Operation“ vybrány skupiny s největším počtem sledujících (ke dni 5.12.2022). U těchto kanálů byl předpoklad že budou orientovány prorusky.

SpecialQForces – 112 479 sledujících

MoD Russia – 87 277 sledujících

Další skupiny byly vybrány na základě klíče „Ukraine Free“, opět ke dni 5.12.2022. U těchto kanálů byl předpoklad proukrajinské orientace.

Donbas Devushka – 49 183 sledujících

Slavyanograd – 42 212 sledujících

Dále byly do výběru přidány tři kanály spojené s novinami a informačními servery.

Breaking News | Times – 56 748 sledujících

The Kyiv Independent – 54 742 sledujících

World News [Breaking News] – 25 511 sledujících

A jako poslední byly vybrány dva malé kanály zaměřené na přeposílání příspěvků z jiných stránek, novin, kanálů a skupin.

The True Story – 142 sledujících

Ukraine – Russia War – 32 sledujících

### **1.7.3 Sběr dat z Telegramu**

Základem vytvořeného Python scriptu je knihovna Telethon (58), dovolující komunikaci s API telegramu. K využití této knihovny je nutno nejprve vytvořit Telegramový účet a získat potřebné přihlašovací klíče „api\_id“ a „api\_hash“.

Tyto údaje jsou ve scriptu použity k navázání kontaktu s Telegramovým klientem. Dále je před samotným sběrem nutné definovat kanály skrz které se příspěvky sbírají a vytvořit struktury pro uchování získaných dat.



Obrázek 8 - Script pro sběr dat z Telegramu – inicializace (Zdroj: vlastní)

```
# Import knihoven
from telethon.sync import TelegramClient
import pandas as pd
import datetime
from telethon.errors import ChannelPrivateError

# Přihlasovací údaje do telegram API
api_id = '*****'
api_hash = '*****'
phone = '*****'

# Definování jmen kanálů
groups = ['WorldNews', 'live_news_times',
          'theverytruestory', 'latestukraine', 'KyivIndependent_official',
          'SpecialQForces', 'mod_russia_en', 'DonbassDevushka', 'Slavyangrad']

# Navázání spojení
client = TelegramClient('test', api_id, api_hash)
client.connect()

# Vytvoření struktur pro uchování získaných dat
df = pd.DataFrame()
forwards_list = []
forwards_df = pd.DataFrame(
    forwards_list,
    columns=[
        "From",
        "Username",
    ],
)
```

Po úspěšném přihlášení skript iteruje skrze všechny definované kanály a sbírá posledních tisíc zpráv (ke dni 5.12.2022) Pokud zpráva obsahuje „forward“, tedy pokud byla původně napsána pro jiný kanál nebo skupinu, skript si zaznamená její název. Po průchodu se data uloží do excelu.

Obrázek 9 - Script pro sběr dat z Telegramu - sběr dat (Zdroj: vlastní)

```
# Hledání skrz všechny skupiny
for group in groups:
    with TelegramClient(None, api_id, api_hash) as client:
        # Vybrání posledních tisíc zpráv od aktuálního data.
        for message in client.iter_messages(group, 1000, reverse=False):
            print(message)

            # Získání informací o datu, hastagu, autorovi textu a forwardnutí příspěvku
            data = {"group": group, "sender": message.sender_id, "text": message.text, "date": message.date,
                    "forward": message.forward}

            # Uložení dat
            temp_df = pd.DataFrame(data, index=[1])
            df = df.append(temp_df)

            # Pokud se jedná o forwardnutou zprávu, uloží se z jakého kanálu byla forwardnuta
            if message.forward is not None:
                try:
                    f_from_id = message.forward.original_fwd.from_id
                    if f_from_id is not None:
                        ent = client.get_entity(f_from_id)
                        username = ent.username
                        forwards_df = pd.DataFrame(
                            forwards_list,
                            columns=[
                                "From",
                                "Username",
                            ],
                        )
                        forwards_list.append(
                            [
                                group,
                                username,
                            ]
                        )
                except Exception as e:
                    if e is ChannelPrivateError:
                        print("Private channel")
                    continue

# Uložení získaných dat
forwards_df.to_excel("C:\\Users\\jirih\\Desktop\\Dataseť\\Telegram\\forwards.xlsx", index=False)
df['date'] = df['date'].dt.tz_localize(None)
client.disconnect()
df.to_excel("C:\\Users\\jirih\\Desktop\\Dataseť\\Telegram\\data_{}.xlsx".format(datetime.date.today()), index=False)
```

První spuštění skriptu bylo s ručně vybranými kanály a z každého kanálu sesbíral posledních 1000 zpráv. Tento průchod vygeneroval datovou sadu 9000 zpráv a seznam 97 dalších kanálů, které byly v této datové sadě zmíněny pomocí funkce „forward“.

Druhý průchod byl s náhodným vzorkem nově získaných kanálů.

Konkrétně se jedná o:

„nytimes\_world“ zmíněno v „World News [Breaking News]“

„truesummaryworld“ zmíněno v „The True Story“

„ukr\_leaks\_eng“ zmíněno v „Slavyangrad“

„ghostnewsx“ zmíněno v „Slavyangrad“

„rybar“ zmíněno v „DonbassDevushka“

„NewResistance“ zmíněno v „DonbassDevushka“

Druhý průchod tak vygeneroval dalších 5158 zpráv a 131 souvisejících kanálů. Celkově bylo vygenerováno 14158 příspěvků publikovaných mezi 27. listopadem a 5. prosincem.

## 1.8 Začištění dat

Surová získaná data obsahují znaky a části textu, které nejsou pro strojová učení vhodná. Především se jedná o emotikony, URL odkazy, reklamní bannery, označení jiných uživatelů anebo hashtagy. Z tohoto důvodu bylo před samotným využitím získaných dat přistoupeno k jejich začištění pomocí Python scriptu a ručního začištění.

Některé příspěvky jsou odsazeny na několik řádků. Častokrát je mezi jednotlivými částmi několika řádková mezera, původně sloužící ke grafické úpravě příspěvku. Tyto části textu není nutné začišťovat, jelikož transformační model bílé znaky ignoruje.

### 1.8.1 Začištění dat z Twitteru

Za pomoci skriptu se ze všech příspěvků odstranili emotikony, URL odkazy a hashtagy. K tomuto automatickému odstranění byla využita knihovna „cleantext“ (59).

*Obrázek 10 - Script pro začištění Telegramových dat (Zdroj: vlastní)*

```
# Import knihoven
import pandas as pd
from cleantext import clean
import re

# Načtení surových dat
df = pd.read_excel(r"C:\Users\jirih\Desktop\Dulezite\22b.PostprodukceTwitter\Raw.xlsx")
df['NEW TEXT'] = 'Placeholder'

# Průchod skrz všechny příspěvky
for ind in df.index:
    df['NEW TEXT'][ind] = ((re.sub(r"?:\@|\(http?:://|http?:://|https?:://|\(https?:://|www|#\S+", "", clean(df["text"][ind], no_emoji=True))))

# Uložení začištěných dat
df.to_excel("C:\Users\jirih\Desktop\Dulezite\22b.PostprodukceTwitter\Zpracovana.xlsx")
```

Dále proběhla náhodná ruční kontrola, která neodhalila žádné příspěvky, které by byly nevhodně nebo zcela prázdné. Výsledný začištěný datový soubor obsahuje 15600 příspěvků.

### 1.8.2 Začištění dat z Telegramu

Sebraná data z telegramu vykazují velké znečištění především odkazy, hashtagy, emotikony a fotografiemi. Pomocí skriptu byly z dat odstraněny všechny URL odkazy, emotikony, nápisy označující odkaz, znaky „(, , ,)“, „ „, „[“, „]“ a „\_“, které se na Telegramu ve velké míře využívají ke grafickému ohraničení příspěvků a nemají žádný informační význam, a podpisy jednotlivých kanálů.

K odstranění emotikonů a URL adres se využila knihovna „cleantext“ (59), další úpravy proběhly nahrazením závadného textu prázdným řetězcem.

Obrázek 11 - Script pro začištění Telegramových dat (Zdroj: vlastní)

```
# Import knihoven
import pandas as pd
from cleantext import clean
import re

# Načtení surových dat
df = pd.read_excel("C:\\Users\\jirih\\Desktop\\Dulezite\\22a.PostProdukceTelegram\\Raw.xlsx")
df['NEW TEXT'] = 'Placeholder'

# Průchod skrz všechny příspěvky
for ind in df.index:
    # Odstranění nežádoucích elementů
    df['NEW TEXT'][ind] = ((re.sub(r"(?:@|\\(http?:|https?:|\\(https?:|www|#)S+", "", clean(df["text"][ind], no_emoji=True)))
        .replace("*", "") #Odstranění znaku *
        .replace("[read full]", "") #odkaz na clanek
        .replace("[A]", "") #odkaz na clanek
        .replace("[", "") #Odstranění znaku [
        .replace("]", "") #Odstranění znaku ]
        .replace("(", "") #Odstranění znaku (
        .replace(")", "") #Odstranění znaku )
        .replace("_", "") #Odstranění znaku _
        .replace('Join Slavyangrad chat. Your opinion matters.', "") #Odstranění reklamy
        .replace('join slavyangrad', "") #Odstranění reklamy
        .replace('join slg intelligence briefings, strategy and analysis, expert community', "") #Odstranění reklamy
    )

# Uložení začištěných dat
df.to_excel("C:\\Users\\jirih\\Desktop\\Dulezite\\22a.PostProdukceTelegram\\Zpracovana.xlsx")
```

Po proběhnutí automatického skriptu proběhla náhodná ruční kontrola získaných dat. Z té se zjistilo, že data z kanálu „rybar“ byla v drtivé většině psaná v ruštině a nesplňují tak stanovené podmínky. Z tohoto důvodu byla ručně odstraněna.

Dále se v souboru objevily prázdné příspěvky a příspěvky s textem „NaN“ Prázdné příspěvky vznikly z důvodu, že původní text příspěvku obsahoval pouze URL adresu, odkaz nebo hashtag a byl zcela vymazán čistícím skriptem. Hodnotu „NaN“ obsahovaly ty příspěvky, jež byly prázdné před spuštěním začist'ovacího skriptu, tedy ty příspěvky, které obsahovaly pouze multimediální soubor, v převážné většině fotografii.

Ručně byly vymazány všechny příspěvky z definovaných tří kategorií. Finální začištěný datový soubor obsahuje 10655 příspěvků.

## 1.9 Porovnání získaných začištěných datových sad

Finální datové sady obsahují 15600 příspěvků získaných z Twitteru a 10655 příspěvků z Telegramu. Obě datové sady byly porovnány pomocí textového analyzátoru QuitaUP (60). Další analýza proběhla pomocí napsaného skriptu na porovnání délek jednotlivých příspěvků. Přehled všech porovnávaných údajů je zobrazen v tabulkách „Statistiky z porovnání datových sad“, „Přehled délek slov v datových sadách“ a „Deset nejčastějších významových slov v datových sadách“.

Ze získaných dat vyplývá, že příspěvky na Telegramu jsou v průměru o 26,1 slov delší a mají tak v průměru téměř dvojnásobnou délku než příspěvky na Twitteru. Slovní zásoba, tedy počet unikátních slov v datovém souboru je také větší na Telegramu, kde dosahuje 27072 unikátních slov. Na Twitteru je to pouze 19582 slov. V tomto počtu jsou zahrnuty všechny slova, tedy i člen, citoslovce, překlady a vymyšlená slova. Poměr unikátních slov na celkovém objemu slov je opět lehce vyšší u Telegramu, kde činí 0,04575, na Twitteru je tato hodnota 0,04264. Entropii, tedy míru diverzity slovníku a tím pádem bohatší slovník má s hodnotou 10,43 opět větší Telegram. Twitter v tomto měření dosahuje hodnoty 9,62. Analýza délky jednotlivých slov ukázala podobné rozložení na obou platformách, s tím že Telegram vykazoval větší výskyt delších slov.

Při porovnání deseti nejčastějších významových slov, tedy slov, které nesou informační význam (tabulka „Deset nejčastějších významových slov v datových sadách“ záměrně vynechává slova typu předložka, zájmena, citoslovce, členy, slovesa být a mýt a všechna slova ne-nesoucí informační význam.) se dle předpokladu na Twitteru vysoko umístily slova, jež byla použita při sběru dat, „Russia“, „Ukraine“, „Ukraine“, „War“, „Russian“ a „Putin“. Jediné klíčové slovo „Zelensky“ se nedostalo mezi deset nejčastějších. V datové sadě bylo zastoupeno pouze 507x (0,086 %). Kromě těchto slov byly nejčastěji zastoupeny slova „People“, „Nato“, „Against“ a „World“. Na Telegramu bylo v deseti nejčastějších slovech šest stejných jako na Twitteru, a to „Russia“, „Ukraine“, „Ukraine“, „People“, „Russian“ a „War“. Kromě nich se nejčastěji objevovali slova „Military“, „Forces“, „Said“ a „Region“.

Tabulka 4 - Statistiky z porovnání datových sad (Zdroj: vlastní)

	Datová sada z Telegramu		Datová sada z Twitteru	
Počet získaných příspěvků	10656		15601	
Celkově slov	Absolutně	Průměrně	Absolutně	Průměrně
	591802	55,5369745	459283	29,43933081
Počet unikátních slov	Absolutně	Průměrně	Absolutně	Průměrně
	27072	2,54054054	19582	1,25517595
Poměr unikátních slov na celkovém objemu slov	0,04574503		0,042636022	
Entropie	10,430529		9,621172	



Tabulka 5 - Přehled délek slov v datových sadách (Zdroj: vlastní)

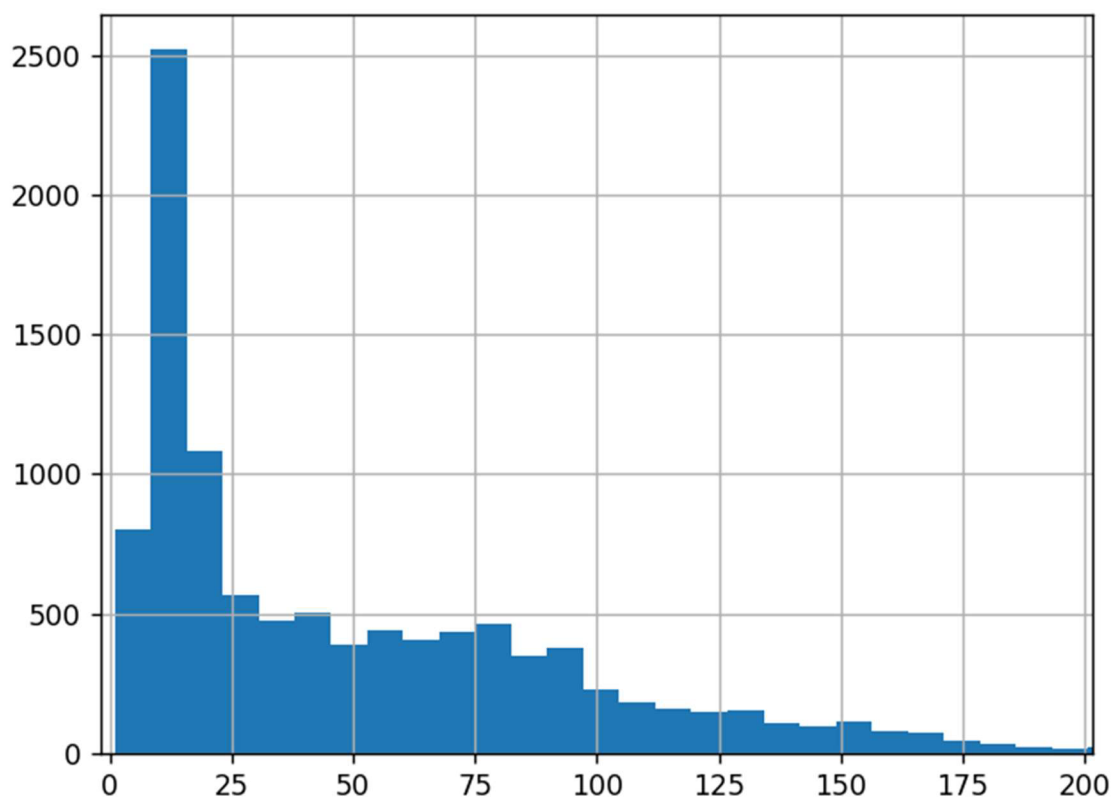
Délka slova	Datová sada z Telegramu	Datová sada z Twitteru
1	20282	20218
2	88435	75275
3	101281	98467
4	81390	71422
5	60522	46834
6	57869	43763
7	64240	47811
8	45394	22005
9	35296	15701
10	18053	9225
11	9266	4238
12	4551	2419
13	2462	1069
14	1922	577
15+	839	259

Tabulka 6 - Deset nejčastějších významových slov v datových sadách (Zdroj: vlastní)

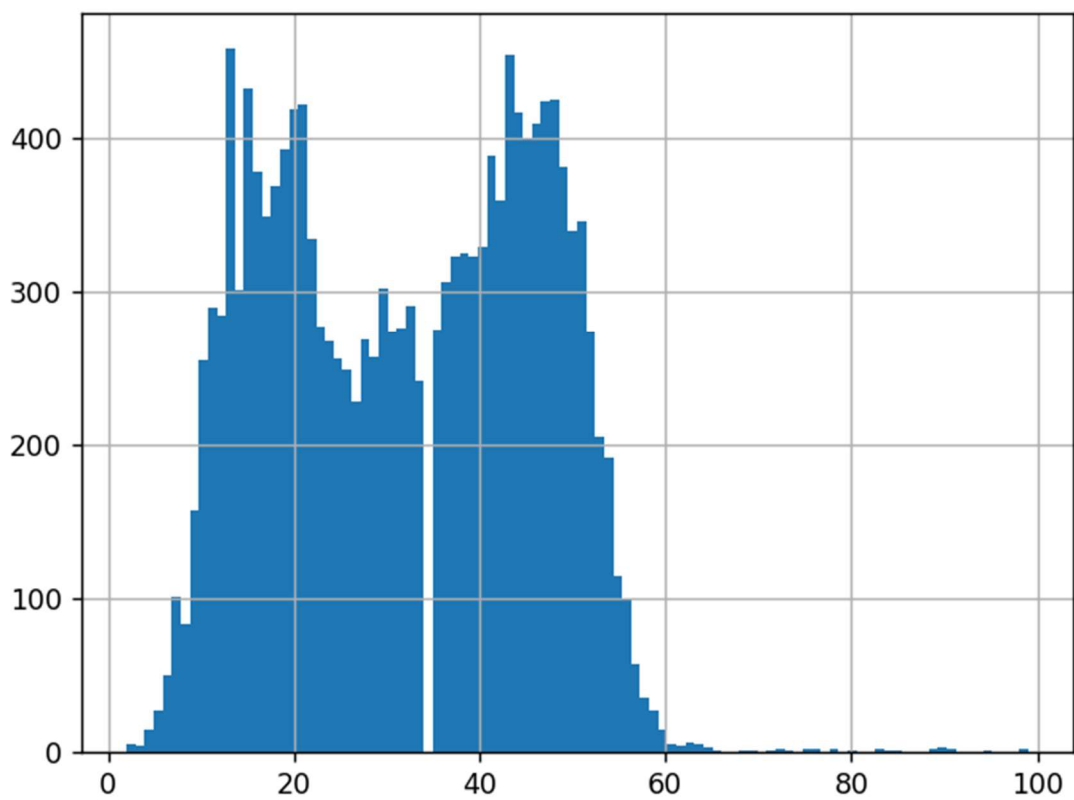
Pořadí	Datová sada z Telegramu			Datová sada z Twitteru		
	Slovo	Absolutní hodnota výskytu	%	Slovo	Absolutní hodnota výskytu	%
1.	ukraine	4 754	0,8	war	15 282	3,33
2.	russian	4 256	0,72	ukraine	12 001	2,61
3.	ukrainian	2 706	0,46	russia	8 407	1,83
4.	russia	2 587	0,44	russian	3 449	0,75
5.	military	2 263	0,38	putin	2 619	0,57
6.	forces	1 911	0,32	people	1 287	0,28
7.	said	1 768	0,3	nato	1 122	0,24
8.	people	1 694	0,29	against	1 119	0,24
9.	region	1 247	0,21	ukrainian	1 102	0,24
10.	war	1 133	0,19	world	1 030	0,22

Následující tabulky ukazují graf rozložení délky slov v jednotlivých datových sadách. Osa x označuje délku příspěvku, konkrétně počet slov. Osa y vyznačuje počet příspěvků.

Graf 1 - Rozložení příspěvku na Telegramu podle délky (Zdroj: vlastní)



Graf 2 - Rozložení příspěvku na Twitteru podle délky (Zdroj: vlastní)



## 1.10 Analýza pomocí transformačního modelu BERT

K práci s natrénovaným transformačním modelem BERT byla využita knihovna „transformers“ (61), kód využitý v napsaném skriptu byl inspirován kódem zveřejněným na stránkách serveru TowardsDataScience (62)

Celý skript je rozdělen do dvou částí, první vyžaduje označená data, na kterých se BERT během stanoveného počtu epoch finálně doladí. Druhá načte vzorky ze získaných datových sad a pomocí natrénovaného modelu je ohodnotí.

### 1.10.1 Označení dat pro doladění modelu BERT

Ke správnému finálnímu doladění je potřeba poskytnout transformačnímu modelu vzorová označená data podle kterých se model bude učit. V této práci byly zvoleny dva přístupy. Ruční označení náhodného vzorku získaných dat z Twitteru a Telegramu a použití části již označených dat z datové sady PTC (63).

Ruční označení probíhalo v několika krocích, zaprvé se z obou získaných datových sad náhodně vybralo 150 příspěvků a na základě tabulky zmíněné v sekci 3.3 byly označeny buď jako propagandistické nebo ne-propagandistické. Následně bylo ručně vybráno dalších 50 příspěvků, tak aby se celkový poměr propagandistických a ne-propagandistických dat rovnal přibližně 1:1. Posledním krokem bylo přidání 17-ti vzorových propagandistických textů, získaných z práce „Classification of Propaganda on Fragment Level: Using Logistic Regression with Handcrafted Contextual Features“ (64) a 9-ti ne-propagandistických textů získaných z článku na zpravodajském webu Aljazeera. (65) Tímto postupem vznikla označená sada o velikosti 226 příspěvcích ve složení 107 ne-propagandistických a 119 propagandistických příspěvků.

Datová sada PTC byla získána ve formě jednotlivých článků, rozdělených na segmenty, a separátních textových dokumentů, které v jednotlivých člancích pomocí pozice znaků identifikovali propagandistické sekvence. Aby bylo možno tyto data použít, byl napsán skript, který prošel všemi články, rozdělil je po jednotlivých sekvencích a tyto sekvence označil jako propagandistické, pokud se v nich vyskytoval propagandistický segment. Takto vznikla označená sada s 10028 ne-propagandistickými a 3518 propagandistickými segmenty.

Aby byla sada vyvážená, z ne-propagandistických se náhodně vybralo 3518 segmentů. Ve spojení s 3518 segmenty propagandistickými vytvořili finální označenou sadu



o velikosti 7036. Při jednotlivých testování se velikost této sady upravovala, ale vždy tak aby poměr propagandistických a ne propagandistických segmentů byl přibližně 1:1.

### 1.10.2 Finální doladění modelu BERT

Prvním krokem v samotném testování bylo naučit natrénovaný BERT model na detekci propagandy podle vytvořených označených dat.

Nejprve se označená data tokenizují, tedy jednotlivým slovům nebo části slovům, takzvaným „Tokens“, jsou přiřazeny jejich identifikační čísla „Tokens ID“ a je určeno jestli se dané části při trénování mají brát v úvahu „Attention Mask“. Také se v tomto kroku přiřadí speciální tokeny na začátek a konec jednotlivých segmentů/příspěvků. [CLS] token je vždy přiřazen na začátek a transformační model podle něj pozná, kde segment začíná. Podobně token [SEP] označuje ukončení segmentu. Také se na volné nevyužité pozice doplní token [PAD], tak aby finální velikost všech sekvencí odpovídala definované délce a byla u všech stejná.

Obrázek 12 - Script BERT – Tokenizace (Zdroj: vlastní)

```
# Tokenizace
def preprocessing(input_text, tokenizer):
    return tokenizer.encode_plus(
        input_text,
        add_special_tokens = True,
        max_length = 512,
        pad_to_max_length = True,
        return_attention_mask = True,
        return_tensors = 'pt'
    )

for sample in text:
    encoding_dict = preprocessing(sample, tokenizer)
    token_id.append(encoding_dict['input_ids'])
    attention_masks.append(encoding_dict['attention_mask'])

token_id = torch.cat(token_id, dim = 0)
attention_masks = torch.cat(attention_masks, dim = 0)
labels = torch.tensor(labels)
```

Tabulka 7 - Příklad tokenizace (Zdroj: vlastní)

Tokens	Tokens ID	Attention Mask
[CLS]	101	1
macro	26632	1
##n	2078	1
says	2758	1
new	1047	1
security	3036	1
architecture	4294	1
should	2323	1
give	2507	1
guarantees	21586	1
for	2005	1
russia	3607	1
[SEP]	102	1
[PAD]	0	0
[PAD]	0	0
[PAD]	0	0

V dalším kroku se data rozdělí do dvou skupin, jedna určena k trénování a druhá k ověření, jak dobře je model schopen rozpoznávat zadaný úkol. Nastavení poměru těchto dvou skupin bylo stanoveno na hodnotu 0,2 ve prospěch trénovacích dat.

Obrázek 13 - Script BERT - Rozdělení dat (Zdroj: vlastní)

```
# Rozdělení dat na trénovací a ověřovací skupinu
train_idx, val_idx = train_test_split(
    np.arange(len(labels)),
    test_size = 0.2,
    shuffle = True,
    stratify = labels)

train_set = TensorDataset(token_id[train_idx],
                           attention_masks[train_idx],
                           labels[train_idx])

val_set = TensorDataset(token_id[val_idx],
                        attention_masks[val_idx],
                        labels[val_idx])
```

Následně je nutné specifikovat základní parametry potřebné pro finální doladění BERT modelu, velikost dávkování dat (Batch size), křivku učení (Learning rate) a počet

průchodů (epochs). Tyto parametry byly stanoveny podle doporučení pro nastavení modelu BERT při finálním doladění, jenž vzešlo ze studie „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding“ (36). Ta doporučuje nastavit dávkování dat na hodnotu 32 nebo 16, v této práci se používala hodnota 16. Křivku učení na hodnoty 5e-5, 3e-5 nebo 2e-5, v této práci se používala hodnota 5e-5. Počet epoch byl doporučen na 2, 3 nebo 4, ve většině případů se v této práci používaly epochy tři, v případě testování jiného množství epoch je ve výsledcích uvedena poznámka upřesňující, s jakou hodnotou tohoto parametru byl test proveden.

Po specifikaci všech nutných parametrů je možno spustit finální doladění modelu. Během něj se nejprve postupně načítají trénovací data a posléze se vyhodnocuje výkonost na ověřovacích datech. K tomuto byly ve skriptu použity části kódů zveřejněných na stránkách serveru TowardsDataScience (62). Ty kromě samotného procesu trénování a ověření také dovolují sledovat u každé epochy pět proměnných označujících výkonost trénovaného modelu.

Hlavními parametry, které se u modelu sledují jsou:

- **Ztráta při trénování** (Training loss) - poměr chyb modelu na celkové množině testovaných dat.
- **Přesnost** (Accuracy) – Poměr správně vyhodnocených úryvků na celkovém počtu úryvků.
- **Přesnost** (Precision) – Poměr správně vyhodnocených propagandistických úryvků na všech úryvcích označených jako propagandistické.
- **Rozpoznání** (Recall) – Poměr správně vyhodnocených propagandistických úryvků na všechny propagandistické úryvky.
- **Specifičnost** (Specificity) - Poměr správně vyhodnocených ne propagandistických úryvků na všechny ne propagandistické úryvky.

### 1.10.3 Vyhodnocení získaných datových sad

Poté co je model finálně doladěn je možné ho využít k samotné detekci propagandistických příspěvků. Každý model byl vždy otestován na třech datových sadách. Na datech, která sloužila k natrénování modelu a na datových sadách nebo jejich částech získaných z Twitteru a Telegramu.

Při vyhodnocování byly kromě finální predikce zaznamenány i údaje, jak moc se model přiklonil k jednotlivé variantě. Tyto údaje se dají využít pro získání informace, jak moc si je model „jistý“ svou predikcí.

Obrázek 14 - Skript BERT - Ukázka predikce, (Zdroj: vlastní)

```
# Načtení dat a přidání sloupců pro predikci a sílu jednotlivých variant
df2 = pd.read_excel(r"C:\Users\jirih\Desktop\Dulezite\3.BERT\Twitter.xlsx")
df2['Propaganda'] = 'Placeholder'
df2['Neni'] = 'Placeholder'
df2['Je'] = 'Placeholder'

# Průchod všemi příspěvky
for ind in df2.index:
    test_ids = []
    test_attention_mask = []
    sentenc = df2['text'][ind]

    # Tokenizace příspěvků
    encoding = preprocessing(sentenc, tokenizer)
    test_ids.append(encoding['input_ids'])
    test_attention_mask.append(encoding['attention_mask'])
    test_ids = torch.cat(test_ids, dim = 0)
    test_attention_mask = torch.cat(test_attention_mask, dim = 0)

    # Předpověď
    with torch.no_grad():
        output = model(test_ids.to(device), token_type_ids = None, attention_mask = test_attention_mask.to(device))

    # Zapsání výsledků
    prediction = 1 if np.argmax(output.logits.cpu().numpy()).flatten().item() == 1 else 0
    df2['Propaganda'][ind] = prediction
    df2['Je'][ind] = output.logits.cpu().numpy()[0][0]
    df2['Neni'][ind] = output.logits.cpu().numpy()[0][1]

# Uložení výsledků
df2.to_excel("C:\\Users\\jirih\\Desktop\\Dulezite\\3.BERT\\TwitterSPredikci.xlsx")
```

## 5. Výsledky a diskuse

### 1.11 Výsledky měření

Dotrénování modelu BERT a predikce výsledků proběhla celkem šestkrát, vždy za použití odlišných vstupních parametrů. Dvakrát proběhlo dotrénování na vlastních označených datech, jednou při nastavení dvou epoch a jednou při třech epochách. Čtyřikrát proběhlo dotrénování za využití označené datové sady PTC, při každém průchodu byly zvolené odlišné velikosti poskytnutých označených dat, jejichž výběr probíhal náhodně.

Samotná predikce byla u každého modelu spuštěna na různě velkých datových sadách. Nejmenší model využil 500 vzorků, největší všechna získaná data. V případě použití pouze části dat, byla tato data náhodně vybrána. Jednotlivé nastavení u každého průchodu je vypsáno v tabulce „Vstupní parametry při jednotlivých průchodech“.

Tabulka 8 - Vstupní parametry při jednotlivých průchodech (Zdroj: vlastní)

Průchody	Označená data			Telegram	Twitter	Nastavení modelu
	Způsob získání	Velikost datové sady	% propagandistických vzorků	Velikost datové sady		Počet epoch
1.	Vlastní	228	47,14	3000	3000	2
2.	Vlastní	228	47,14	3500	3500	3
3.	PTC	7036	49,97	500	500	1
4.	PTC	200	50	5000	5000	3
5.	PTC	2250	49,31	5000	5000	3
6.	PTC	7036	49,97	10655	15600	4

#### 1.11.1 Získaná data

Jednotlivé průchody vykazovaly rozdílné výsledky přesnosti měření. Přehled výkonosti jednotlivých modelů je vypsán v tabulce „Přehled výkonosti modelů“. Údaje jsou vždy vázány k poslední epoše.

Tabulka 9 - Přehled výkonosti modelů (Zdroj: vlastní)

Průchody	Ztráta při trénování (v %)	Přesnost (v %)	Preciznost (v %)	Rozpoznání (v %)	Specifičnost (v %)
1.	45,77	69,35	65,63	83,80	61,18
2.	29,61	58,04	56,55	59,52	55,56
3.	50,46	79,58	89,15	67,01	91,04
4.	7,98	91,68	96,30	87,78	94,44
5.	13,64	78,02	81,44	72,60	83,10
6.	6,37	80,82	79,15	84,31	77,56

Natrénované modely byly znovu ověřeny na označených datech. Proběhlo srovnání mezi původními hodnotami a hodnotami určenými dotrénovaným modelem BERT. Také bylo provedeno porovnání pouze těch příspěvků, u kterých rozdíl mezi váhou jednotlivých variant byl v absolutní hodnotě větší než 1. Výsledky jsou zaznamenány v tabulce „Ověření modelů na označených datech“

Tabulka 10 - Ověření modelů na označených datech (Zdroj: vlastní)

Průchody	Označená data			
	Původní zastoupení propagandistických příspěvků (%)	Modelem vyhodnocené propagandistické příspěvky (%)	Špatně vyhodnocené příspěvky (%)	Špatně vyhodnocené příspěvky, které měli rozdíl mezi váhami jednotlivých variant větší než 1 (%)
1.	47,14	52,42	10,57	6,82
2.	47,14	48,02	22,90	3,88
3.	49,97	41,05	11,82	5,16
4.	50,00	49,00	2,00	2,01
5.	49,31	47,80	5,15	3,58
6.	49,97	50,71	4,26	3,45

Části získaných datových sad, v případě 6. průchodu všechna získaná data, byly otestovány jednotlivými dotrénovanými modely. U každého průchodu bylo zaznamenáno procentuální zastoupení propagandistických příspěvků, jak v celém vybraném datovém souboru, tak v části definované těmi příspěvky, u kterých model stanovil rozdíl mezi vahami jednotlivých variant větší než 1. Naměřené hodnoty jsou zaznamenány v tabulce „Tabulka procentuálního zastoupení propagandistických příspěvků podle dotrénovaného modelu“



Tabulka 11 - Tabulka procentuálního zastoupení propagandistických příspěvků podle dotrénovaného modelu (Zdroj: vlastní)

Průchody	Telegram		Twitter	
	Celé testované data (%)	Příspěvky u kterých byl rozdíl mezi variantami větší než 1 (%)	Celé testované data (%)	Příspěvky u kterých byl rozdíl mezi variantami větší než 1 (%)
1.	<b>57,20</b>	46,92	<b>58,28</b>	54,04
2.	<b>57,38</b>	25,96	<b>51,47</b>	26,72
3.	<b>65,33</b>	70,08	<b>84,61</b>	91,22
4.	<b>97,78</b>	98,60	<b>99,10</b>	99,58
5.	<b>57,59</b>	59,01	<b>83,80</b>	88,08
6.	<b>70,99</b>	73,90	<b>92,58</b>	94,77

### 1.11.2 Statistická analýza získaných dat

Pro statistickou analýzu byla využita analýza kategoriálních data to na datech získaných během 6. průchodu. Jako nejkvalitnější sice vyšel 4. průchod, který dosahoval 91,68 % přesnosti a při kontrolním ohodnocování špatně ohodnotil pouze 2 % zpráv. Vzhledem k malému rozsahu označených dat byl ale nakonec vybrán právě nejrobustnější 6. průchod. Ten dosáhl druhého nejlepšího výsledku. Přesnost vykazoval 80,82 % a špatně označených příspěvků v rámci testovacího ohodnocení bylo 4,26 %.

K ověření získaných dat byla sestavena asociační tabulka. Vzhledem k dostatečné velikosti sebraných dat byl využit  $\chi^2$  – test. K výpočtům byl využit statistický software SAS.

(66)

Tabulka 12 - Asociační tabulka (Zdroj: vlastní)

Propagandistický příspěvek	Sociální síť	
	Twitter	Telegram
Ano	14442	7564
Ne	1158	3091

Nulová hypotéza zní: Na základě dat získaných při 6. průchodu, výskyt propagandy nezávisí na sociální síti.

Alternativní hypotéza: Výsledek 6. průchodu potvrdil, že výskyt propagandy závisí na sociální síti.

Tabulka 13 - Výsledek  $\chi^2$  – testu (Zdroj: vlastní, na základě (66))

<b>Výsledek <math>\chi^2</math> – testu</b>	2174,8929
<b>Hladina pravděpodobnosti</b>	<0,0001
<b><math>\Phi</math> koeficient</b>	0,2878
<b>Kontingenční koeficient</b>	0,2766
<b>Křížový poměr</b>	5,0964

Závěrem je že při hladině významnosti 0,05 se hypotéza  $H_0$  zamítá a hypotéza  $H_1$  se přijímá. Tedy  $\chi^2$  – test potvrdil, že výskyt propagandy závisí na konkrétní sociální síti. Na základě dat získaných modelem BERT při 6. průchodu, test určil že šance na výskytu propagandy na sociální síti Twitter je 5,0964 krát větší než na sociální síti Telegram.  $\Phi$  koeficient vyšel 0,2878 a jedná se tedy o slabou závislost.

## 1.12 Diskuse

Výsledky ukázaly že v rámci shromážděných datových sad, vykazuje datová sada z Twitteru větší koncentraci propagandistických příspěvků, než je tomu u dat získaných z platformy Telegram. Pět ze šesti dotrénovaných modelů BERT označili v testovaných datech z Twitteru více propagandistických zpráv. Pouze jeden model označil více propagandistických příspěvků na Telegramu. Zároveň se jednalo o model s nejmenší naměřenou přesností, 58,04 % a při kontrolním ohodnocování označených dat vybral 22,95 % dat nesprávně, což je nejvíce ze všech testovaných modelů.

Označená data se ukázala jako velice důležitá a zásadní, modely dotrénované s ručně označenými daty vykazovali menší přesnost při identifikování propagandy. Při ověřování během samotného dotrénování modelu dosahovali přesnosti 69,35 % a 58,04 %, při kontrolním průchodu tyto modely chybovali v 10,57 % a 22,90 % případů. Naproti tomu modely využívající datový soubor PTC vykazovaly přesnost 79,58 %, 91,68 %, 78,02 % a 80,82 % a chybovaly v 11,82 %, 2 %, 5,15 % a 4,26 %.

Ukázalo se, že malý rozsah označených dat může vést k rozdílnému dotrénování modelu při jednotlivých průchodech. Jak 1. průchod tak 2. průchod měli na vstupu stejná označená data, ale jejich přesnost se lišila o 11,31 %. Velikost hraje roli i v kvalitě detekce, jak ukázal 4. průchod, který dosahoval nejlepších výsledků, ale zároveň ohodnocoval



příspěvky velice striktně. Na Twitteru označil 99,10 % a na Telegramu 97,78 %. V porovnání s ním průchody 3., 5. a 6., dotréované na stejné, ale rozšířené sadě, ohodnotili jako propagandistické 84,61 %, 83,80 % a 92,58 % Twitterových a 65,33 %, 57,59 % a 70,99 % Telegramových příspěvků. Přes rozdílné výsledky jednotlivých modelů, je zajímavé, že v drtivé většině bylo na Twitteru označeno více příspěvků jako propagandistických než na Telegramu.

Projevení různého počtu trénovacích epoch závisí na použitých označených datech, u ručně označených se počet epoch prakticky neprojevil. U sady PTC bylo pozorováno výrazné zlepšení s každou další epochou. Toho si lze všimnout na porovnání 3. a 6. průchodu, které byly trénované na shodné sadě, ale rozdílném množství epoch. Zatímco 3. průchod se vytrévoval pouze na 1 epoše, 6. průchod na 4 epochách. Naměřená přesnost byla u 3. průchodu 79,58 % a kontrolní označení vyhodnotilo špatně 11,82 % příspěvků. 6. průchod dosáhl přesnosti 80,82 % a kontrolní označení chybovali v 4,26 % případech.

Za zmínku stojí, že pokud omezíme modely, tak aby kromě přiřazování hodnot „ano“ a „ne“ mohli označovat data jako „nejasná“ a nastavíme určitou hranici, která musí být překročena, chybovost jednotlivých modelů klesá. U 1., 2., 3., 5. a 6. průchodu se chybovost výrazně snížila, u 4. průchodu se ze 2 % zanedbatelně zvedla na 2,01 %. Pokud bychom toto omezení aplikovali na datových souborech z Twitteru a Telegramu tak se nám výsledky upraví v závislosti na použité označené sadě. U ručně označené se procento propagandistických příspěvků zmenší a u 2. průchodu, jenž jako jediný označoval u Telegramu více propagandistických příspěvků se hodnoty tak měř vyrovnají, 25,96 % ku 26,72 %. U sady PTC se při zavedení tohoto omezení výskyt propagandy vždy zvedl, nikdy se ovšem nezměnilo pořadí jednotlivých platforem.

Provedené měření tedy ukázala že pomocí transformačního modelu BERT lze určit jaká sociální síť vykazuje větší míru propagandistických příspěvků.

### **1.12.1 Porovnání výsledků s jinými studii**

Přesnost doladěných modelů je možno nepřímo srovnat s výsledky soutěží „NLP4IF-2019“ a „SemEval 2020“ (30) (31). Nutno ovšem podotknout, že tyto soutěže se zaměřovali na mnohem složitější úkol, a to rozpoznávání přesných propagandistických úryvků v textu. Naproti tomu, tato práce si kladla za cíl pouze určit je-li daný příspěvek propagandistický či nikoli.

U „SemEval 2020“ dosáhl nejlepší model 66,37 % přesnosti a v průměru se nejlepší modely pohybovaly kolem 60 %. U „NLP4IF-2019“ dosáhl nejlepší model 67,24 % a nejlepší modely se pohybovaly opět přes 60 %. Modely použité v této práci vykazovaly v jednodušší úloze přesnost od 58,04 % po 91,68 %. Toto srovnání nelze brát jako přímé, jelikož zadání úlohy, testovací data i rozsah modelu byly odlišné. Může ovšem sloužit k porovnání toho, jakou přesnost transformační modely v obtížných úkolech mají.

Porovnání ostatních výsledků, jako je četnost propagandy na Twitteru a Telegramu, nemohou být provedeny, jelikož v době psaní této práce nebyly nalezeny žádné výzkumné práce, které by se zabývaly přímo tímto problémem. Existují práce zaměřené na charakteristiku propagandy na Twitteru (67), nebo konkrétní události jako například Covid-19 (68), žádné z nich ovšem neanalyzují propagandu na Twitteru nebo Telegramu jako celek a neporovnávají jejich četnost.

### **1.12.2 Omezení zvoleného přístupu**

Během práce na tomto výzkumu vyšlo najevo mnoho omezení, které bylo nutné překonat, nebo obejít.

Nejprve bylo nutno vyřešit jak si při získávání datových souborů poradit s rozdílným fungováním sociálních sítí Twitter a Telegram, tak aby získané datové sady byly možná co nejpodobnější. Byl zvolen přístup filtrace na základě konkrétní události, zkombinované s klíčovými slovy a frázemi a konkrétním použitým jazykem, angličtinou. Ačkoli se tento přístup zdá jako správný, přinesl deformaci dat, která mohla konečně výsledky ovlivnit.

Další překážkou je formát příspěvků, ty často obsahují multimediální soubory, odkazy, emotikony a znaky. Převážně na Telegramu jsou ke grafické úpravě příspěvku často využívány znaky typu „/“, „(“, „{“ a „[„. V této práci se tento problém vyřešil odstraněním všech informací přímo nesouvisejících s napsanou zprávou, tento zásah opět deformoval získaná data a mohl mít vliv na konečné výsledky.

Velkým omezením se ukázala označená data vhodná k dotrénování, jak odhalila jednotlivá měření, jsou tyto data a jejich správnost velice důležité při dotrénování modelu. Zde použité způsoby, tedy ruční označení a datový soubor PTC, splnili účel, ale mohly ovlivnit výsledky. Ručně označený datový soubor nebyl dostatečně velký a konzistentní a sada PTC se nevztahovala přímo k testované události.

Omezením, přirozeně vyplívající z rozsahu této práce, je i velikost samotného trénovaného modelu BERT. V této práci byly k dotrénování využívány relativně malé datové sady, v řádech stovek, maximálně tisíců. Pro přesnější a spolehlivější závěry, by bylo nutno robustnější testování.

### **1.12.3 Možné rozšíření využití metodiky**

Využitý postup nabízí spoustu možností pro vylepšení a rozšíření. Jak již bylo popsáno výše, během výzkumu bylo nutno potýkat se se spoustou omezeními a překážkami. Ty v každém bodě nabízejí příležitosti pro zlepšení zde použitého postupu.

Jako první se nabízí shromáždění robustnější datové sady. U Twitteru by bylo vhodné zaměřit se na širší výběr klíčových slov, zahrnout hashtagy a především neomezovat se jen na anglicky psané příspěvky. U Telegramu by se rozsáhlejší datové sady dalo docílit širším výběrem počátečních kanálů a vícenásobným vyhledáváním skrz „forwardnuté“ kanály. I tady by zcela jistě pomohlo začlenit do výběru i jinak než anglicky psané příspěvky.

Vylepšení by bylo možné i u začištění získaných dat. Například kvalitnější filtrace, tak aby jednotlivé části byly odstraněny opravdu pouze v případě, že nenesou žádnou informaci. U dotrénování modelu BERT by se dala používat širší datová sada ručně označených příspěvků. Samotné dotrénování pak nabízí možnost trénovat skrz více epoch.

Jedná-li se o změny které by mohli být na použitém postupu provedeny, jeví se jako hlavní využití více transformačních modelů. Rozdělení jednotlivých příspěvků na menší sekce a posuzovat každou sekci samostatně. Nebo opakované posouzení, kde se model učí na datech, které sám ohodnotil.

Širokým prostorem k zaplnění je oblast automatizace celého procesu vyhodnocování, kde by s dostatečně velkým označeným datovým souborem, byl prostor posuzovat příspěvky okamžitě při zveřejňování. Takovéto příspěvky by následně mohli být označeny jako potenciálně rizikové.

## 6. Závěr

Hlavním cílem bylo porovnat četnost výskytu propagandy na sociálních sítích Twitter a Telegram. Výzkum propagandy se ukázal jako nesmírně složitý a multioborový problém. Samotná definice toho, co to propaganda vlastně je, jak jí definujeme a co všechno pod tento neurčitý pojem spadá, je velice náročná. Úkol ručního určení, je-li příspěvek propagandistický či nikoli, se ukázal obtížný a oblast využití počítačové technologie k takovému úkolu dělá teprve své první dětské krůčky.

Prvním dílčím cílem bylo identifikovat vhodnou metodu k detekci propagandy na sociálních sítích Twitter a Telegram. Při výběru vhodné metody se vycházelo z pokroku, který byl učiněn v posledních několika letech. Na základě výsledků několika soutěží se ukázaly transformační modely, v čele s modelem BERT a RoBERTa a jejich různé modifikace, velice slibné při porozumění a pochopení přirozeného jazyka. Velikou výhodou těchto modelů je jejich relativní snadnost a rychlost používání. Tyto vlastnosti se potvrdili i během této práce, kde se použitý transformační model BERT osvědčil.

Druhým dílčím cílem bylo aplikovat vybranou metodu na datových souborech získaných z Twitteru a Telegramu. Sběr dat proběhl, díky nativním API jednotlivých platform a široké škále volně dostupných python knihoven, bezproblémově. Větší oříšek představovalo, jak definovat vhodný filtr pro získání správných dat. Zde se ukázalo, že použitá metoda zachytila správný směr, ale nebyla dostatečně robustní. Podobné lze říci i u začišťování získaných dat.

Práce s modelem BERT za použití knihovny „transformers“ nepředstavovala větší problém. Knihovna dovoluje specifikovat širokou škálu nastavení při dotrénování modelu. Vzhledem k tomu, že byl použit nejrozšířenější, základní transformační model, existují k dispozici četné ukázky jeho použití. Drobným zádrhelem byla časová náročnost, kdy jednotlivé průchody trvaly desítky hodin a v případě špatně definovaných parametrů bylo nutno celý proces začít od znovu.

Provedené měření ukázala, že v získaných datových sadách, propaganda převažuje na straně Twitteru. Toto zjištění by se dalo se přisoudit tomu, jak je definována propaganda. Jak bylo ukázáno během této práce, v nemalé míře je často spojena s citově zabarveným vyjadřováním. Platforma Twitter je založena na publikování svých dojmů a názorů ve větší míře, než je tomu u Telegramu, který více slouží k šíření obecných informací. Otázkou

zůstává, jak zjištěné výsledky ovlivnila absence rusky a ukrajinsky psaných příspěvků, či jiné z omezení zmíněných v podkapitole „5.2.2 Omezení zvoleného přístupu“.

Právě kvůli mnohým ústupkům, které se v této práci musely udělat, nelze získaná data vztahovat na celý Twitter a Telegram, ale pouze na získané datové sady. K zjištění reálného zastoupení propagandy na jednotlivých sítích by bylo zapotřebí vytvořit několik robustních modelů, každý zaměřen na jiný aspekt propagandy. Dále by bylo potřeba shromáždit širší a větší datové sady, na kterých by se testovali její jednotlivé stránky. I tak zůstává otázkou, jestli by takto navržený model byl schopen obsáhnout a identifikovat všechny tváře propagandy.

## 7. Citovaná literatura

- (1) DEMILT, Jessica. The Origins of Twitter. In: *Pennington Creative* [online]. Internet: Pennington Creative, 2022 [cit. 2022-08-03]. Dostupné z: <https://penningtoncreative.com/the-origins-of-twitter/>
- (2) DIXON, S. Most popular social networks worldwide as of January 2022, ranked by number of monthly active users. In: *Statista* [online]. Internet: Statista Inc., 2022 [cit. 2022-08-03]. Dostupné z: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- (3) VAROL, Onur, Emilio FERRARA, Clayton A. DAVIS, Filippo MENCZER a Alessandro FLAMMINI. Online Human-Bot Interactions: Detection, Estimation, and Characterization. *Computer Science* [online]. 2017, **2017**(1), 6 [cit. 2022-08-18]. Dostupné z: <http://arxiv.org/abs/1703.03107>
- (4) YAQUB, M. How Many Tweets per Day 2022. In: *Renolon* [online]. Internet: Renolon, 2022 [cit. 2022-08-18]. Dostupné z: <https://www.renolon.com/number-of-tweets-per-day>
- (5) GLOBAL SOCIAL MEDIA STATISTICS. In: *Data Portal* [online]. Internet: Kepios, 2022 [cit. 2022-08-18]. Dostupné z: <https://datareportal.com/social-media-users>
- (6) BAUMGARTNER, Jason, Savvas ZANNETTOU, Megan SQUIRE a Jeremy BLACKBURN. The Pushshift Telegram Dataset. *Proceedings of the International AAAI Conference on Web and Social Media* [online]. 2020, **2020**(14), 840-847 [cit. 2022-08-18]. Dostupné z: <https://arxiv.org/abs/2001.08438>
- (7) SPROULE, J. Michael a English ERIC CLEARINGHOUSE ON READING. *Channels of Propaganda* [online]. 1994 [cit. 2022-08-18]. ISBN 9780927516617. ISSN ERICRIE0.
- (8) NAIMUSHINOVA, Anastasiia. *Analýza mediálního obrazu války na Ukrajině ve vybraných médiích*. Smetanovo nábřeží 6, 2021. Diplomová práce. UNIVERZITA KARLOVA FAKULTA SOCIÁLNÍCH VĚD.
- (9) ELLUL, Jacques. Propaganda: The Formation of Men's Attitudes. In: *Propaganda: The Formation of Men's Attitudes* [online]. 1. Internet: archive.org, 2015, s. 79-84 [cit. 2022-07-25]. ISBN 9780394718743. Dostupné z: [https://archive.org/details/Propaganda\\_201512/page/n9/mode/2up](https://archive.org/details/Propaganda_201512/page/n9/mode/2up)
- (10) FARKAS, Johan a Christina NEUMAYER. Disguised Propaganda from Digital to Social Media. *Second International Handbook of Internet Research* [online]. 2018, 1-19 [cit. 2022-08-18]. Dostupné z: doi:10.1007.978-94-024-1202-4\_33-1
- (11) AUERBACH, J a R CATRONOVO. *The Oxford Handbook of Propaganda Studies* [online]. 2014, 1-19 [cit. 2022-08-18]. ISSN edsair. Dostupné z: doi:10.1093/oxfordhb/9780199764419.001.0001
- (12) KELLNER, Ansgar, Lisa RANGOSCH, Christian WRESSNEGGER a Konrad RIECK. *Political Elections Under (Social) Fire? Analysis and Detection of Propaganda on Twitter* [online]. 2019 [cit. 2022-08-18].
- (13) BRADSHAW, S a P HOWARD. Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation. *Computational Propaganda Research Project* [online]. 2017 [cit. 2022-08-05]. ISSN edsair.

- (14) ABOKHODAIR, Norah, Daisy YOO a David W. MCDONALD. *Dissecting a Social Botnet: Growth, Content and Influence in Twitter* [online]. 2016 [cit. 2022-08-18]. Dostupné z: doi:10.1145/2675133.2675208
- (15) CHAUDHARI, Deptii a Ambika PAWAR. A Systematic Comparison of Machine Learning and NLP Techniques to Unveil Propaganda in Social Media. *Journal of Information Technology Research* [online]. 2021, **15**(1), 1-14 [cit. 2022-08-18]. ISSN 19387857. Dostupné z: doi:10.4018/JITR.299384
- (16) TONG, Yanqi a Shaohua LEI. War of Position and Microblogging in China. *Journal of Contemporary China* [online]. 2013, **22**(80), 292-311 [cit. 2022-08-18]. ISSN 10670564. Dostupné z: doi:10.1080/10670564.2012.734084
- (17) ARO, Jessikka. The cyberspace war: propaganda and trolling as warfare tools. *European View* [online]. 2016, **15**(1), 121-132 [cit. 2022-08-18]. ISSN 17816858. Dostupné z: doi:10.1007/s12290-016-0395-5
- (18) SUŠÁNKOVÁ, Hana. *Automatická detekce dezinformací*. Arna Nováka 1, 2021. Bakalářská práce. Masarykova Univerzita Filozofická fakulta. Vedoucí práce RNDr. Michal Černý.
- (19) KHANDAY, Akib Mohi Ud Din, Qamar Rayees KHAN a Syed Tanzeel RABANI. Identifying propaganda from online social networks during COVID-19 using machine learning techniques. *International Journal of Information Technology: An Official Journal of Bharati Vidyapeeth's Institute of Computer Applications and Management* [online]. 2021, **13**(1), 115-122 [cit. 2022-08-18]. ISSN 25112104. Dostupné z: doi:10.1007/s41870-020-00550-5
- (20) TAM, Adrian. A Gentle Introduction to Vector Space Models. In: *Machine Learning Mastery* [online]. Internet, 2021 [cit. 2022-08-18]. Dostupné z: <https://machinelearningmastery.com/a-gentle-introduction-to-vector-space-models/>
- (21) SALTON, Gerard a Michael J MCGILL. *Introduction to Modern Information Retrieval*. 1983.
- (22) GHELANI, Shreya. From Word Embeddings to Pretrained Language Models. In: *Towards Data Science* [online]. 2019 [cit. 2022-08-18]. Dostupné z: <https://towardsdatascience.com/from-word-embeddings-to-pretrained-language-models-a-new-age-in-nlp-part-1-7ed0c7f3dfc5>
- (23) DEERWESTER, S., S. T. DUMAIS, T. K. LANDAUER, G.W. FURNAS a R. A. HARSHMAN. Indexing by latent semantic analysis. *J. Soc. Inf. Sci.* [online]. 1990, **41**(6), 391-407 [cit. 2022-08-18].
- (24) TAZZYMAN, Sam. *Latent Semantic Analysis* [online]. In: . [cit. 2022-08-18]. Dostupné z: <https://moj-analytical-services.github.io/NLP-guidance/LSA.html>
- (25) THOMAS, H. *Probabilistic latent semantic indexing*. 22nd Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, 1999, 50-57.
- (26) EMERGING TECHNOLOGY. King – Man + Woman = Queen: The Marvelous Mathematics of Computational Linguistics. In: *MIT Technology review* [online]. 2015 [cit. 2022-08-18]. Dostupné z: <https://www.technologyreview.com/2015/09/17/166211/king-man-woman-queen-the-marvelous-mathematics-of-computational-linguistics/>
- (27) TOSIN, Adewumi, Liwicki FOTEINI a Liwicki MARCUS. Word2Vec: Optimal hyperparameters and their impact on natural language processing downstream tasks.

- Open Computer Science* [online]. 2022, **12**(1), 134-141 [cit. 2022-08-18]. ISSN 22991093. Dostupné z: doi:10.1515/comp-2022-0236
- (28) MIKOLOV, Tomas, Kai CHEN, Greg CORRADO a Jeffrey DEAN. *Efficient Estimation of Word Representations in Vector Space* [online]. 2013 [cit. 2022-08-18].
- (29) PENNINGTON, J., R. SOCHER a Christopher MANNING. *Glove: Global Vectors for Word Representation* [online]. 2014 [cit. 2022-08-18]. ISSN edsair. Dostupné z: doi:10.3115/v1/d14-1162
- (30) MARTINO, G. Da San, A. BARRÓN-CEDEÑO, H. WACHSMUTH, R. PETROV a P. NAKOV. *SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles* [online]. 2020 [cit. 2022-08-12].
- (31) MARTINO, Giovanni Da San, Alberto BARRÓN-CEDEÑO a Preslav NAKOV. *Findings of the NLP4IF-2019 Shared Task on Fine-Grained Propaganda Detection* [online]. 2019 [cit. 2022-08-18].
- (32) PETERS, Matthew E., Mark NEUMANN, Mohit IYYER, Matt GARDNER, Christopher CLARK, Kenton LEE a Luke ZETTLEMOYER. *Deep contextualized word representations* [online]. 2018 [cit. 2022-08-18].
- (33) HOWARD, Jeremy a Sebastian RUDER. *Universal Language Model Fine-tuning for Text Classification* [online]. 2018 [cit. 2022-08-18].
- (34) VASWANI, Ashish, Noam SHAZEER, Niki PARMAR, Jakob USZKOREIT, Llion JONES, Aidan N. GOMEZ, Lukasz KAISER a Illia POLOSUKHIN. *Attention Is All You Need* [online]. 2017 [cit. 2022-08-18].
- (35) PHI, Michael. Illustrated Guide to Transformers- Step by Step Explanation. In: *Towards Data Science* [online]. 2020 [cit. 2022-08-18]. Dostupné z: <https://towardsdatascience.com/illustrated-guide-to-transformers-step-by-step-explanation-f74876522bc0>
- (36) DEVLIN, Jacob, Ming-wei CHANG, Kenton LEE a Kristina TOUTANOVA. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* [online]. 2018 [cit. 2022-08-18].
- (37) TAYLOR, Wilson L. “Cloze Procedure”: A New Tool for Measuring Readability. *Journalism* [online]. 1953, **30**(4), 415-433 [cit. 2022-08-18]. ISSN 10776990. Dostupné z: doi:10.1177/107769905303000401
- (38) LIU, Yinhan, Myle OTT, Naman GOYAL et al. *RoBERTa: A Robustly Optimized BERT Pretraining Approach* [online]. 2019 [cit. 2022-08-18].
- (39) MARTINO, Giovanni Da San, Stefano CRESCI, Alberto BARRON-CEDENO, Seunghak YU, Roberto DI PIETRO a Preslav NAKOV. *A Survey on Computational Propaganda Detection* [online]. 2020 [cit. 2022-08-18].
- (40) RASHKIN, Hannah, Eunsol CHOI, Jin Yea JANG, Svitlana VOLKOVA a Yejin CHOI. *Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking* [online]. 2017 [cit. 2022-08-18]. ISSN edsair. Dostupné z: doi:10.18653/v1/d17-1317
- (41) BARRON-CEDENO, A., I. JARADAT a P. NAKOV. *Propopy: Organizing the news based on their propagandistic content* [online]. 2019 [cit. 2022-08-18]. ISSN edsair. Dostupné z: doi:10.1016/j.ipm.2019.03.005
- (42) MARTINO, Giovanni Da San, Seunghak YU, Alberto BARRÓN-CEDEÑO, Rostislav PETROV a Preslav NAKOV. *Fine-Grained Analysis of Propaganda in News Articles* [online]. 2019 [cit. 2022-08-18].



- (43) YOOSUF, Shehel a Yin YANG. Fine-Grained Propaganda Detection with Fine-Tuned BERT. *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda* [online]. 2019 [cit. 2022-08-18]. ISSN edsair. Dostupné z: doi:10.18653/v1/d19-5011
- (44) VLAD, George-alexandru, Mircea-adrian TANASE, Cristian ONOSE a Dumitru-clementin CERCEL. Sentence-Level Propaganda Detection in News Articles with Transfer Learning and BERT-BiLSTM-Capsule Model. *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda* [online]. 2019 [cit. 2022-08-18]. ISSN edsair. Dostupné z: doi:10.18653/v1/d19-5022
- (45) RAVIKIRAN, Manikandan, Amin Ekant MULJIBHAI, Toshinori MIYOSHI, Hiroaki OZAKI, Yuta KOREEDA a Sakata MASAYUKI. *Hitachi at SemEval-2020 Task 12: Offensive Language Identification with Noisy Labels using Statistical Sampling and Post-Processing* [online]. 2020 [cit. 2022-08-18].
- (46) JURKIEWICZ, Dawid, Łukasz BORCHMANN, Izabela KOSMALA a Filip GRALIŃSKI. *ApplicaAI at SemEval-2020 Task 11: On RoBERTa-CRF, Span CLS and Whether Self-Training Helps Them* [online]. 2020 [cit. 2022-08-18].
- (47) CHERNYAVSKIY, Anton, Dmitry ILVOVSKY a Preslav NAKOV. *Aschern at SemEval-2020 Task 11: It Takes Three to Tango* [online]. 2020 [cit. 2022-08-18].
- (48) WHITE, Anna, Sumeet DUA a Radhika MEDURY. Divisive Language and Propaganda Detection using Multi-head Attention Transformers with Deep Learning BERT-based Language Models for Binary Classification. *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda* [online]. 2019 [cit. 2022-08-18]. ISSN edsair. Dostupné z: doi:10.18653/v1/d19-5014
- (49) MADABUSHI, Harish Tayyar, Elena KOCHKINA a Michael CASTELLE. *Cost-Sensitive BERT for Generalisable Sentence Classification with Imbalanced Data* [online]. 2020 [cit. 2022-08-18].
- (50) GUPTA, Pankaj, Khushbu SAXENA, Usama YASEEN, Thomas RUNKLER a Hinrich SCHÜTZE. *Neural Architectures for Fine-Grained Propaganda Detection in News* [online]. 2019 [cit. 2022-08-18].
- (51) TWITTER. *Search Tweets* [online]. In: . 2022 [cit. 2022-08-18]. Dostupné z: <https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query>
- (52) DATA SCIENCE PARICHAY. *Get Data From Twitter API v2 in Python*. In: *Data Science Parichay* [online]. 2022 [cit. 2022-08-18]. Dostupné z: <https://datascienceparichay.com/article/python-get-data-from-twitter-api-v2/>
- (53) BAUMGARTNER, Jason Michael a Savvas ZANNETTOU. *Pushshift Telegram Ingest API* [online]. In: . 2019 [cit. 2022-08-18]. Dostupné z: <https://github.com/pushshift/telegram>
- (54) PYTHON. *Telethon 1.24.0*. In: *Python Software Foundation* [online]. 2021 [cit. 2022-08-10]. Dostupné z: <https://pypi.org/project/Telethon/>
- (55) Snsrape. In: *GitHub* [online]. github: github, 2022 [cit. 2022-12-05]. Dostupné z: <https://github.com/JustAnotherArchivist/snsrape>

- (56) Twitter dokumentace. In: *Twitter* [online]. online: Twitter, 2022 [cit. 2023-01-06]. Dostupné z: <https://developer.twitter.com/en/docs/twitter-api/v1/rules-and-filtering/search-operators>
- (57) TGStat. In: *TGStat* [online]. TGStat, 2022 [cit. 2022-12-05]. Dostupné z: <https://tgstat.com/ratings/channels>
- (58) Telethon Dokumentace. In: *Telethon* [online]. Telethon, 2022 [cit. 2023-12-05]. Dostupné z: <https://docs.telethon.dev/en/stable/modules/client.html>
- (59) CleanText. In: *Pypi* [online]. [cit. 2022-12-12]. Dostupné z: <https://pypi.org/project/cleantext/>
- (60) ČECH, Radek, Vladimír MATLACH a Miroslav KUBÁT. *QuitaUp* [online]. In: . [cit. 2023-01-07]. Dostupné z: <https://korpus.cz/quitaup/>
- (61) Transformers. In: *Pypi* [online]. Pypi [cit. 2022-11-12]. Dostupné z: <https://pypi.org/project/transformers/>
- (62) ALBANESE, Nicolo Cosimo. Fine-Tuning BERT for Text Classification. In: *Towards Data Science* [online]. 2022: Towards Data Science [cit. 2023-01-04]. Dostupné z: <https://towardsdatascience.com/fine-tuning-bert-for-text-classification-54e7df642894>
- (63) *Propaganda Analysis Project* [online]. In: . [cit. 2022-10-09]. Dostupné z: <https://propaganda.math.unipd.it/fine-grained-propaganda-emnlp.html>
- (64) MAISSAN, Sylvain. *Classification of Propaganda on Fragment Level: Using Logistic Regression with Handcrafted Contextual Features* [online]. 2020 [cit. 2022-12-07]. Dostupné z: <https://studenttheses.uu.nl/bitstream/handle/20.500.12932/36498/Scriptie.docx.pdf?sequence=1>. University Utrecht. Vedoucí práce Dong Nguyen.
- (65) *Aljazeera* [online]. 2022 [cit. 2022-12-10]. Dostupné z: <https://www.aljazeera.com/>
- (66) SAS INSTITUTE INC. *SAS 9.4* [online]. [cit. 2022-02-24]. Dostupné z: [https://www.sas.com/cs\\_cz/home.html](https://www.sas.com/cs_cz/home.html).
- (67) GUARINO, Stefano, Noemi TRINO, Alessandro CELESTINI, Alessandro CHESSA a Gianni RIOTTA. *Characterizing networks of propaganda on twitter: a case study* [online]. 2020, 23 [cit. 2023-01-07]. Dostupné z: [https://www.researchgate.net/publication/344143008\\_Characterizing\\_networks\\_of\\_propaganda\\_on\\_twitter\\_a\\_case\\_study/link/5fc483af299bf104cf947be7/download](https://www.researchgate.net/publication/344143008_Characterizing_networks_of_propaganda_on_twitter_a_case_study/link/5fc483af299bf104cf947be7/download)
- (68) KUSWANTI, Ana. Trump's Twitter Propaganda During Covid-19. *Jurnal the Messenger* [online]. 2021, **13**(3), 223-237 [cit. 2022-12-08]. ISSN 20861559. Dostupné z: doi:10.26623/themessenger.v13i3.3991

## 8. Seznam obrázků, tabulek, grafů a zkratk

### 1.13 Seznam obrázků

Obrázek 1 - Vizualizace modelu Bag-Of-Words (Zdroj: (22)) .....	16
Obrázek 2 - Vzorec pro výpočet TF (Zdroj: (23)) .....	17
Obrázek 3 - Vzorec pro výpočet IDF (Zdroj: (23)) .....	17
Obrázek 4 - Diagram fungování CBOW a Skip-gram typů modelu Word2Vec (Zdroj: (28)) .....	19
Obrázek 5 - Architektura Transformačních modelů (Zdroj: (35)).....	22
Obrázek 6 - Diagram postupu využitý týmem ApplicaAI (Zdroj: (47)).....	29
Obrázek 7 - Script pro sběr dat z Twitteru (Zdroj: vlastní) .....	34
Obrázek 8 - Script pro sběr dat z Telegramu – inicializace (Zdroj: vlastní) .....	37
Obrázek 9 - Script pro sběr dat z Telegramu - sběr dat (Zdroj: vlastní).....	38
Obrázek 10 - Script pro začištění Telegramových dat (Zdroj: vlastní) .....	39
Obrázek 11 - Script pro začištění Telegramových dat (Zdroj: vlastní) .....	40
Obrázek 12 - Script BERT – Tokenizace (Zdroj: vlastní) .....	45
Obrázek 13 - Script BERT - Rozdělení dat (Zdroj: vlastní).....	46
Obrázek 14 - Script BERT - Ukázka predikce, (Zdroj: vlastní).....	48

### 1.14 Seznam tabulek

Tabulka 1 - Druhy propagandistické techniky (Zdroj: (43), vlastní překlad).....	25
Tabulka 2 – Přehled použitých technik na SemEval 2020 (Zdroj: (31)).....	27
Tabulka 3 - Přehled verzí využitých transformátorů týmem Hitachi (Zdroj: (46)).....	28
Tabulka 4 - Statistiky z porovnání datových sad (Zdroj: vlastní).....	41
Tabulka 5 - Přehled délek slov v datových sadách (Zdroj: vlastní).....	42
Tabulka 6 - Deset nejčastějších významových slov v datových sadách (Zdroj: vlastní) ....	42
Tabulka 7 - Příklad tokenizace (Zdroj: vlastní) .....	46
Tabulka 8 - Vstupní parametry při jednotlivých průchodech (Zdroj: vlastní).....	49
Tabulka 9 - Přehled výkonosti modelů (Zdroj: vlastní).....	50
Tabulka 10 - Ověření modelů na označených datech (Zdroj: vlastní).....	50
Tabulka 11 - Tabulka procentuálního zastoupení propagandistických příspěvků podle dotrénovaného modelu (Zdroj: vlastní) .....	51
Tabulka 12 - Asociační tabulka (Zdroj: vlastní).....	51
Tabulka 13 - Výsledek $\chi^2$ – testu (Zdroj: vlastní, na základě (66)).....	52

### 1.15 Seznam grafů

Graf 1 - Rozložení příspěvku na Telegramu podle délky (Zdroj: vlastní).....	43
Graf 2 - Rozložení příspěvku na Twitteru podle délky (Zdroj: vlastní).....	43

## 9. Přílohy

Jednotlivé Python skripty využity ke sběru datových sad z Twitteru a Telegramu, k jejich začištění a k samotné práci s transformačním modelem BERT, jsou v celém rozsahu k dispozici na adrese „<https://github.com/Pericul/BakalrskaPrace>“.