

UNIVERZITA KARLOVA

FAKULTA SOCIÁLNÍCH VĚD

Institut politologických studií

Katedra politologie

Bakalářská práce

2024

Adam Rušíkvas

UNIVERZITA KARLOVA

FAKULTA SOCIÁLNÍCH VĚD

Institut politologických studií

Katedra politologie

**Rozpoznávání proevropského a protievropského
diskurzu v českém mediálním prostoru**

Bakalářská práce

Autor práce: Adam Rušíkvas

Studijní program: Politologie a mezinárodní vztahy

Vedoucí práce: Dr. rer. pol. Michal Parížek, M.Sc., Ph.D.

Rok obhajoby: 2024

Prohlášení

1. Prohlašuji, že jsem předkládanou práci zpracoval samostatně a použil jen uvedené prameny a literaturu.
2. Prohlašuji, že práce nebyla využita k získání jiného titulu.
3. Souhlasím s tím, aby práce byla zpřístupněna pro studijní a výzkumné účely.

Ve Vrchlabí dne 26. 05. 2024

Adam Rušíkvas

Bibliografický záznam

RUŠIKVAS, Adam, 2024. *Rozpoznávání proevropského a protievropského diskurzu v českém mediálním prostoru*. Bakalářská práce. Praha: Univerzita Karlova, Fakulta sociálních věd, Institut politologických studií, Katedra politologie. Vedoucí bakalářské práce Dr. rer. pol. Michal Parížek, M.Sc., Ph.D.

Rozsah práce: 67 969 znaků

Poděkování

Děkuji vedoucímu této práce Dr. rer. pol. Michalovi Parížkovi, M.Sc., Ph.D. za drahocenné rady, připomínky a velkou trpělivost. Dále bych také rád poděkoval své rodině a přítelkyni, kteří mi dodávali potřebnou podporu.

Abstrakt

Tato bakalářská práce se zabývá rozpoznáváním proevropského a protievropského diskurzu v českém mediálním prostoru. Právě média mohou totiž prostřednictvím specifických rámců a sentimentů ovlivňovat postoje veřejnosti k evropské integraci. Hlavním cílem této práce je vytvořit klasifikační model založený na metodách strojového učení s učitelem (supervised machine learning, SML), který dokáže přesně a spolehlivě identifikovat sentiment a tematický rámec v českých mediálních textech o Evropské unii. Pro potřeby práce došlo v první řadě k vytvoření trénovacího datasetu obsahující články, které byly dle kódovacích schémat zpracovány tradiční obsahovou analýzou. Následně došlo k ověření výkonnosti samotného klasifikačního modelu, který byl zároveň porovnáván s jednodušší metodou "bag of words". Tato práce upozorňuje na potenciál využívání modelů založených na metodách strojového učení s učitelem, které mohou nahradit zdlouhavé tradiční metody analýz a přispět k bližšímu pochopení evropského diskurzu v českém mediálním prostředí.

Abstract

This bachelor thesis deals with the recognition of pro-European and anti-European discourse in the Czech media space. It is the media that can influence public attitudes towards European integration through specific frames and sentiments. The main goal of this thesis is to develop a classification model based on supervised machine learning method that can accurately and reliably identify sentiment and thematic frames in Czech media texts about the European Union. For the purpose of this work, first of all, a training dataset containing articles that were processed according to coding schemes by traditional content analysis was created. Subsequently, the performance of the classification model itself was verified and compared with the simpler "bag of words" method. This work highlights the potential of using models based on machine learning methods with a teacher, which can replace lengthy traditional analysis methods and contribute to a closer understanding of European discourse in the Czech media space.

Klíčová slova

Evropský diskurz, Evropská integrace, Český mediální prostor, Rámce, Sentiment, Obsahová analýza, Strojového učení, Klasifikační model

Keywords

European discourse, European integration, Czech media space, frames, Sentiment, Content analysis, Supervised machine learning, Classification model

Title

Recognizing pro-European and anti-European discourse in Czech media space

Obsah

Úvod	1
I. Teoretická část	2
1 Evropská integrace a veřejné mínění.....	2
1.1 Vývoj vztahu mezi evropskou integrací a veřejným míněním.....	2
1.2 Faktory ovlivňující postoje k EU a evropské integraci	4
1.3 Shrnutí kapitoly Evropská integraci a veřejném mínění	6
2 Mediální pokrytí Evropské unie	7
2.1 Role médií v demokratické společnosti	7
2.2 Mediální pokrytí EU a jeho význam	8
2.3 Shrnutí kapitoly o Mediální pokrytí Evropské unie.....	9
3 Mediální rámcování a sentiment	10
3.1 Koncept mediálního rámcování	10
3.2 Analýza rámců v kontextu EU	11
3.3 Koncept sentimentu v mediální analýze	13
3.4 Vliv sentimentu na veřejné vnímání k EU	14
3.5 Shrnutí kapitoly mediálního rámcování a sentimentu.....	15
II. Empirická část	16
4 Metodologie.....	16
4.1 Vytváření korpusu	16
4.1.2 Proces sběru a filtrace dat	16
4.1.3 Popis výsledného korpusu.....	17
4.2 Obsahová analýza	18
5 Strojové učení.....	26
5.1 Předzpracování textu.....	26
5.2 Transformace textu na numerické vektory pomocí SBERT.....	26
5.3 Transformace textu na numerické vektory pomocí BoW	28
5.4 Výběr a popis modelu strojového učení.....	29
5.5 Implementace SVM modelu s SBERT transformací.....	30
5.6 Evaluace výkonnosti modelu SVM pro SBERT	31
5.7 Implementace SVM modelu s BoW transformací	32
5.8 Evaluace výkonnosti SVM s BoW.....	33
6 Srovnání výsledků SBERT a BoW transformací.....	35
Závěr.....	37
Seznam zdrojů	39
SEZNAM TABULEK	48

SEZNAM GRAFŮ	49
SEZNAM ZKRATEK.....	50

Úvod

Analýza mediálního obsahu hraje v současném společenskovedním výzkumu stále významnější roli. S rostoucí dostupností digitálních textových dat a rozvojem automatizovaných metod jejich zpracování se otevírají nové možnosti, jak zkoumat mediální pokrytí různých témat a jeho vliv na veřejné mínění. Jednou z oblastí, která v posledních letech přitahuje značnou pozornost výzkumníků, je mediální obraz Evropské unie a proces evropské integrace.

Občané při utváření svých názorů na politické otázky do značné míry spoléhají na informace zprostředkované médii. To platí zejména u témat, s nimiž nemají přímou osobní zkušenost, jako je právě Evropská unie a její fungování (Vliegthart et al., 2008). Způsob, jakým média o EU informují, však není neutrální. Volbou specifických rámců a vyjádřením určitých sentimentů mohou ovlivňovat vnímání a postoje veřejnosti k evropské integraci (Krzyżanowski, 2018). Analýza mediálních rámců a sentimentu ve vztahu k EU proto představuje důležitý nástroj pro pochopení dynamiky veřejného mínění a jeho interakce s mediálním prostředím.

Tato práce si klade za cíl přispět k výzkumu mediálního obrazu Evropské unie se specifickým zaměřením na český kontext, který bývá v podobných studiích často opomíjen (Vliegthart et al., 2008; Krzyżanowski, 2018). Hlavním záměrem je vyvinout robustní klasifikační model založený na metodách strojového učení s učitelem (supervised machine learning, SML), který dokáže spolehlivě identifikovat sentiment a tematické rámce související s EU v českých mediálních textech. Specificky se práce zaměřuje na vytvoření modelu, který by předčil tradiční slovníkové metody, jako je bag of words (BoW), v přesnosti klasifikace sentimentu. Tento přístup má potenciál překonat omezení tradičních metod analýzy, jakým je časově náročné ruční kódování dokumentů, čímž poskytne efektivnější a systematičtější náhled na mediální pokrytí evropských témat.

Teoretická část práce se věnuje konceptuálnímu ukotvení výzkumu v oblasti evropské integrace, veřejného mínění a role médií. Dále jsou představeny klíčové pojmy mediálního rámcování a sentimentu, následně jsou diskutovány možnosti jejich analýzy pomocí metod strojového učení. V empirické části je popsán proces vytvoření korpusu mediálních textů, příprava dat, trénování a evaluace klasifikačního modelu. Dále jsou prezentovány výsledky modelu a diskutovány možnosti a limity zvoleného přístupu. Závěr shrnuje hlavní zjištění a nastiňuje potenciál dalšího výzkumu v této oblasti.

I. Teoretická část

Teoretická část práce se zaměřuje na tři hlavní oblasti, které tvoří konceptuální rámec pro analýzu mediálního obrazu Evropské unie v českém kontextu. V kapitole 1 je diskutován vztah mezi evropskou integrací a veřejným míněním, včetně vývoje od "permisivního konsensu" k "omezujícímu nesouhlasu", dále faktory ovlivňující postoje veřejnosti k EU. V kapitole 2 je pozornost věnována roli médií v informování o evropských tématech a jejich vlivu na vnímání EU ze strany občanů. V kapitole 3 jsou představeny koncepty mediálního rámcování a sentimentu jako klíčové nástroje pro analýzu mediálního obsahu. Dále jsou představeny možnosti jejichž zkoumání pomocí metod strojového učení s učitelem.

1 Evropská integrace a veřejné mínění

Evropská unie stojí na prahu nové éry, kdy veřejné mínění hraje klíčovou roli v jejím budoucím směřování. Tato kapitola poskytuje teoretický základ pro analýzu postojů občanů k EU a jejich vlivu na proces evropské integrace.

1.1 Vývoj vztahu mezi evropskou integrací a veřejným míněním

Vztah mezi evropskou integrací a veřejným míněním prošel v průběhu let významným vývojem. V počátcích integračního procesu byla veřejnost do značné míry vyloučena z rozhodování a převládal tzv. "permisivní konsensus", kdy občané v podstatě přenechávali rozhodování o evropské integraci politickým elitám. Tento konsensus vycházel z předpokladu, že evropská integrace je především technokratickým procesem, který nevyžaduje přímé zapojení veřejnosti. Elity tak mohly postupovat v integračním procesu bez významného ohledu na postoje a preference občanů (Hooghe a Marks, 2009).

Situace se však začala měnit v 90. letech 20. století, kdy se evropská integrace stala více politizovanou a kontroverzní otázkou. S přijetím Maastrichtské smlouvy v roce 1992 a následným rozšiřováním kompetencí EU se integrace začala dotýkat stále širšího okruhu oblastí, které přímo ovlivňovaly každodenní život občanů. Tím se zvýšila viditelnost EU a zájem veřejnosti o dění na evropské úrovni. (Hooghe a Marks, 2009)

Jak zdůrazňují Hobolt a Vries (2016), budoucí směřování EU je neoddelitelně spojeno s podporou a postoji jejích občanů. Tento posun od "permissive consensus" k "constraining dissensus" reflektuje rostoucí politizaci evropských témat a zvyšující se význam veřejného mínění v procesu evropské integrace (Hooghe a Marks, 2009).

1.1.1 Od "permisivního konsensu" k "constraining dissensus"

S rostoucí politizací evropských témat se veřejné mínění stalo důležitým faktorem ovlivňujícím proces integrace. Referenda o evropských otázkách, jako bylo například referendum o Maastrichtské smlouvě ve Francii a Dánsku v roce 1992, referendum o Smlouvě z Nice v Irsku v roce 2001 nebo referendum o Lisabonské smlouvě v Irsku v roce 2008, ukázala, že veřejné mínění může mít zásadní dopad na další vývoj integrace (De Vries, 2018). V případě odmítnutí navrhovaných změn v referendu byly vlády nuceny vyjednat výjimky nebo dodatečné záruky, aby získaly souhlas svých občanů.

Tento posun od "permissive consensus" k "constraining dissensus" odráží rostoucí význam veřejného mínění v procesu evropské integrace. Elity již nemohou postupovat bez ohledu na postoje občanů, ale musí brát v úvahu možné reakce veřejnosti a snažit se získat její podporu pro další integrační kroky. To vede k větší opatrnosti a zdrženlivosti při navrhování a prosazování významných změn v institucionální struktuře a fungování EU (Hooghe a Marks, 2009).

1.1.2 Rostoucí politizace evropských témat

S tím, jak se evropská integrace dotýká stále širšího okruhu oblastí, roste také její politizace. Evropská témata se stávají předmětem vnitrostátní politické soutěže a veřejné debaty. Političtí aktéři, jako jsou politické strany, zájmové skupiny a sociální hnutí, se stále více zapojují do diskusí o evropské integraci a snaží se mobilizovat veřejnost ve prospěch svých postojů (Paul Statham a Hans-Jörg Trenz, 2012).

Hooghe a Marks (2007) v této souvislosti hovoří o "politice politizace" evropské integrace. Podle nich se evropská integrace stala předmětem strategické soutěže mezi politickými aktéry, kteří se snaží využít evropská témata k prosazení svých zájmů a získání podpory voličů. Tato soutěž se odehrává jak na národní, tak na evropské úrovni a ovlivňuje podobu a dynamiku integračního procesu.

Politizace evropské integrace má různé podoby a projevy v jednotlivých členských státech. Závisí na faktorech jako je síla euroskeptických stran, postoje hlavních politických aktérů, míra zapojení země do integračního procesu nebo vnímání národních zájmů ve vztahu k EU. Obecně však platí, že s rostoucí politizací evropských témat se zvyšuje význam veřejného mínění a schopnost občanů ovlivňovat směřování evropské integrace (De Wilde a Zürn, 2012).

1.2 Faktory ovlivňující postoje k EU a evropské integraci

Výzkum veřejného mínění o evropské integraci se postupně vyvíjel od základních studií zaměřených na podporu integrace k analýzám širší škály postojů, včetně euroskepticismu (Sánchez-Cuenca, 2000 ; Hooghe, 2007). Klíčovou otázkou zůstává, co formuje různorodé postoje k EU a její integraci. V tomto kontextu se objevují tři hlavní přístupy k vysvětlení podpory EU: utilitární přístupy, identitní motivace a vliv národního politického kontextu (Boomgaarden et al., 2011).

1.2.1 Utilitární přístupy

Utilitární přístup vychází z předpokladu, že lidé posuzují evropskou integraci na základě jejich očekávaných přínosů a nákladů. Tento přístup se zaměřuje na ekonomické faktory, jako je růst HDP, míra nezaměstnanosti nebo inflace. Zkoumá jejich vliv na postoje k EU (Anderson a Reichert, 1995).

Zastánci utilitárního přístupu argumentují, že lidé, kteří mají vyšší příjmy, lepší vzdělání a pracují v odvětvích, která těží z integrace, budou pravděpodobně podporovat členství v EU a další integraci. Naopak lidé s nižšími příjmy a nižším vzděláním, pracující v odvětvích ohrožených konkurencí v rámci jednotného trhu, budou spíše zastávat euroskeptické postoje (Gabel, 1998).

Utilitární přístup byl aplikován také na úrovni států. Podle této logiky by země, které získávají více prostředků z rozpočtu EU, než do něj přispívají, měly vykazovat vyšší míru podpory integrace (Anderson a Kaltenthaler, 1996).

Ačkoli utilitární faktory hrají důležitou roli při formování postojů k EU, nejsou jediným zdrojem podpory nebo opozice vůči integraci. Výzkum ukazuje, že utilitární úvahy jsou často

podmíněny dalšími faktory, jako je národní identita nebo politická orientace (Hooghe a Marks, 2005).

1.2.2 Identitní motivace

Identitní přístup zdůrazňuje roli národní identity a kulturních faktorů při formování postojů k evropské integraci. Vychází z předpokladu, že lidé vnímají EU jako potenciální hrozbu pro svou národní identitu a suverenitu (Carey, 2002).

Podle zastánců tohoto přístupu budou lidé se silnou vazbou na národní identitu a kulturu méně nakloněni evropské integraci, protože v ní spatřují ohrožení národní svébytnosti a tradičních hodnot. Naopak lidé s kosmopolitní orientací a slabší vazbou na národní identitu budou pravděpodobně podporovat integraci jako způsob posílení nadnárodní spolupráce (Hooghe a Marks, 2005).

Identitní motivace se projevují také ve vztahu k imigraci a kulturní rozmanitosti. Lidé, kteří vnímají imigranty a kulturní odlišnosti jako hrozbu, mají tendenci být více euroskeptičtí, protože EU spojují s volným pohybem osob a multikulturalismem (de Vreese a Boomgaarden, 2005).

Výzkumy ukazují, že identitní faktory jsou často silnějším prediktorem postojů k EU než utilitární úvahy (Hooghe a Marks, 2005). Zároveň však platí, že utilitární a identitní motivace se vzájemně nevyklučují a často působí společně při formování postoje jedince k evropské integraci.

1.2.3 Vliv národního politického kontextu

Třetí přístup se zaměřuje na roli národního politického kontextu při formování postojů k EU. Vychází z předpokladu, že lidé často nemají dostatek informací a znalostí o fungování EU, a proto se při vytváření svých postojů spoléhají na vodítka z národní politické scény (Anderson, 1998).

Podle tohoto přístupu budou lidé, kteří podporují vládní strany a jsou spokojeni s fungováním národních institucí, pravděpodobně podporovat také evropskou integraci. Naopak lidé, kteří jsou nespokojeni s národní politikou a podporují opoziční strany, budou spíše zastávat euroskeptické postoje (Sánchez-Cuenca, 2000).

Vliv národního politického kontextu se projevuje také v tom, jak politické strany rámuji evropská témata a jaké postoje k nim zaujímají. Strany mohou využívat evropskou integraci k mobilizaci voličů a prosazování svých zájmů, což ovlivňuje celkovou dynamiku veřejné debaty o EU (Hooghe a Marks, 2007).

Výzkum ukazuje, že vliv národního politického kontextu na postoje k EU se liší mezi jednotlivými členskými státy a v čase. Závisí na faktorech jako je míra polarizace stranického systému, pozice hlavních stran k evropské integraci nebo důvěra v národní instituce (De Vries, 2018).

1.3 Shrnutí kapitoly Evropská integraci a veřejném mínění

Kapitola poskytuje teoretický základ pro pochopení vztahu mezi evropskou integrací a veřejným míněním a faktorů, které tento vztah ovlivňují. Tyto poznatky mohou být důležité pro vývoj modelu pro analýzu mediálních článků a přispět k jeho schopnosti zachytit relevantní aspekty mediálního pokrytí EU. Zároveň mohou přispět k hlubšímu porozumění role médií v procesu evropské integrace a formování postojů veřejnosti k EU.

2 Mediální pokrytí Evropské unie

Tato kapitola se zabývá rolí médií v demokratické společnosti a významem mediálního pokrytí Evropské unie. Kapitola dále shrnuje, jakou funkci media plní a jaký vliv mají na obyvatelstvo.

2.1 Role médií v demokratické společnosti

Média jsou považována za čtvrtou větev moci ve státě, vedle moci zákonodárné, výkonné a soudní, což odráží jejich klíčovou roli v demokratické společnosti. Hlavní funkcí médií je poskytovat občanům informace, nezbytné pro kvalifikované rozhodování o veřejných záležitostech a aktivní zapojení se do demokratického procesu (Strömbäck a Shehata, 2019).

V ideálním případě by média měla plnit roli hlídacího psa demokracie, kontrolovat mocenské elity a upozorňovat na případné zneužívání moci či porušování demokratických principů. Poskytováním platformy pro veřejnou diskusi a pluralitu názorů média přispívají k vytváření informované a angažované veřejnosti, která je schopna kriticky přemýšlet o společenských problémech a podílet se na jejich řešení (Hanitzsch a Vos, 2018).

Způsob, jakým média informují o společenských a politických tématech, však nemusí být neutrální. Média mohou určitým způsobem rámcovat realitu, zdůrazňují některé aspekty a jiné upozadují, čímž ovlivňují, jak lidé daným tématům rozumí a jaké postoje k nim zaujímají (Cacciatore et al., 2016). Více v kapitole 3 Mediální rámcování a sentiment.

2.1.1 Informační funkce médií

Primární úlohou médií je informovat veřejnost o dění ve společnosti a ve světě. Média fungují jako hlavní zdroj informací o politických, ekonomických a sociálních událostech, o aktivitách vlády a dalších veřejných institucí a o názorech a postojích různých aktérů (McCombs et al., 2014). Tím, že média zprostředkovávají informace, které by jinak byly pro většinu občanů nedostupné, umožňují veřejnosti vytvářet si vlastní názory a postoje a činit informovaná rozhodnutí.

Aby média mohla plnit svou informační funkci, musí poskytovat přesné, vyvážené a nestranné zpravodajství. To znamená, že by měla usilovat o objektivitu, ověřovat fakta, dávat prostor různým názorům a pohledům a jasně oddělovat zpravodajství od komentářů

(Salgado, 2018). Jen tak mohou média přispívat k vytváření kvalitní veřejné debaty a posilovat důvěru občanů v demokratický proces.

2.1.2 Vliv médií na veřejné názory a postoje

Média však neplní pouze informační funkci, ale mají také významný vliv na formování veřejných názorů a postojů. Teorie nastolování agendy (agenda-setting) předpokládá, že média ovlivňují to, o čem lidé přemýšlejí a mluví tím, že některým tématům věnují více pozornosti než jiným (McCombs et al., 2014). Média tak určují, která témata jsou považována za důležitá a hodná veřejné diskuse.

Vliv médií na veřejné názory a postoje je však komplexní a závisí na řadě faktorů, jako je důvěryhodnost médií, individuální charakteristiky příjemců, názorové klima ve společnosti nebo konkurence jiných zdrojů informací. Média tedy nejsou všemocná, ale jejich vliv na veřejnou debatu a politický proces nelze podceňovat (Shehata a Strömbäck, 2018).

2.2 Mediální pokrytí EU a jeho význam

V kontextu Evropské unie hrají média klíčovou roli při informování veřejnosti o dění na evropské úrovni a zprostředkování vztahu mezi EU a jejími občany. Vzhledem k tomu, že EU je složitý a mnohvrstevnatý politický systém, který je pro většinu lidí vzdálený a obtížně srozumitelný, jsou média hlavním zdrojem informací o evropské integraci a politikách EU (Andersen et al., 2016).

2.2.1 Média jako zprostředkovatelé informací mezi EU a veřejností

Média fungují jako most mezi EU a veřejností. Zprostředkovávají informace o aktivitách evropských institucí, o rozhodnutích přijímaných na evropské úrovni, o názorech a postojích politických lídrů i o dopadech evropských politik na každodenní život občanů (Walter, 2017). Tím přispívají k informovanosti veřejnosti o EU a k vytváření evropské veřejné sféry, v níž probíhá diskuse o společných evropských tématech.

Aby média mohla efektivně plnit svou zprostředkující roli, musí poskytovat kvalitní a vyvážené informace o EU. To však není snadný úkol vzhledem ke komplexitě evropské

politiky, jazykové a kulturní rozmanitosti členských států, také často technickému charakteru projednávaných témat (Marquart et al., 2019). Média tak čelí výzvě, jak přiblížit EU občanům a učinit ji srozumitelnější a relevantnější pro jejich každodenní život.

2.2 Dopad mediálního pokrytí na vnímání EU

Způsob, jakým média informují o EU, má významný vliv na to, jak veřejnost vnímá evropskou integraci a jaké postoje k ní zaujímá. Výzkumy ukazují, že pozitivní a podpůrné mediální pokrytí EU vede k vyšší podpoře integrace, zatímco negativní a kritické pokrytí posiluje euroskeptické postoje (Brosius et al., 2019).

Dopad mediálního pokrytí na postoje veřejnosti k EU však není přímočarý a závisí na řadě faktorů. Například důvěryhodnost médií, individuální charakteristiky příjemců, jako jsou jejich politická orientace, hodnoty a identita, spolu s celkovým kontextem veřejné debaty o evropské integraci, hrají důležitou roli. (Brosius et al., 2019). Vliv médií na postoje k EU je tak podmíněný a proměnlivý.

Celkově lze říct, že mediální pokrytí EU hraje klíčovou roli při formování informované a angažované evropské veřejnosti. Tím, že média poskytují informace o dění na evropské úrovni, zprostředkovávají vztah mezi EU a občany a ovlivňují veřejné vnímání a postoje k integraci. V neposlední řadě také přispívají k legitimitě a demokratické odpovědnosti evropského projektu. Zároveň však čelí výzvám, jak učinit EU srozumitelnější a relevantnější pro každodenní život občanů.

2.3 Shrnutí kapitoly o Mediální pokrytí Evropské unie

Poznatky z této kapitoly jsou důležité pro hlavní cíl naší práce, který se zaměřuje na vytvoření modelu pro rozpoznávání mediálních článků týkajících se EU a analýzu polarity sentimentu a rámců v těchto člancích pomocí algoritmů strojového učení. Pochopení role médií a faktorů ovlivňujících mediální pokrytí EU poskytuje teoretický základ pro výzkum a zdůrazňuje význam analýzy mediálního obsahu pro porozumění veřejnému mínění a politickému procesu v kontextu evropské integrace.

3 Mediální rámcování a sentiment

Tato kapitola se zaměřuje na dva klíčové koncepty v analýze mediálního pokrytí Evropské unie: mediální rámcování a polarity sentimentu. Rámcování a sentiment jsou důležitými nástroji pro pochopení, jak média prezentují evropská témata a jaký vliv může mít tato prezentace na veřejné mínění a postoje k evropské integraci.

3.1 Koncept mediálního rámcování

3.1.1 Definice a význam rámcování

Mediální rámcování (framing) je proces, při kterém média vybírají a zdůrazňují určité aspekty reality a tím ovlivňují, jak publikum danou problematiku vnímá a interpretuje (Entman, 1993). Rámce fungují jako interpretační schémata, která dávají smysl komplexním událostem a tématům tím, že je zasazují do známého kontextu a přisuzují jim určitý význam (Vreese, 2005). Rámcování je důležitým nástrojem mediální komunikace, protože umožňuje zjednodušit a zpřístupnit složitá témata širokému publiku.

Volbou specifických rámců mohou média ovlivňovat, které aspekty reality jsou považovány za důležité, jaké příčiny a důsledky jsou zdůrazňovány a jaká řešení jsou navrhována (Semetko a Valkenburg, 2000). Rámcování tak může mít významný dopad na formování veřejného mínění a politických preferencí. Jak zdůrazňují (Lecheler a Vreese, 2018), způsob, jakým jsou témata v médiích rámcována, má významný vliv na to, jak je publikum vnímá a jaké postoje k nim zaujímá. Rámce mohou ovlivňovat, které aspekty problému jsou považovány za relevantní, jak jsou hodnoceny příčiny a důsledky a jaká řešení jsou upřednostňována.

3.1.2 Vliv rámcování na interpretaci a postoje veřejnosti

Výzkumy ukazují, že vystavení určitým rámcům může vést ke změnám v postojích a chování jednotlivců, například v ochotě podporovat určité politiky nebo kandidáty (Lecheler a de Vreese, 2017). Účinky rámcování však mohou být podmíněny individuálními charakteristikami příjemců, jako je jejich politická orientace, hodnoty nebo znalosti o daném tématu (Boomgaarden et al., 2013). Rámce, které jsou v souladu s existujícími postoji a

přesvědčeními jedince, mají větší šanci ovlivnit jeho názory než rámce, které jsou s nimi v rozporu.

Vliv rámcování na veřejné mínění a politické preference je obzvláště významný v kontextu evropské integrace, která je komplexním a mnohvrstevnatým procesem. Jak poznamenávají Galpin a Trenz (2019) pozitivní rámcování evropské integrace může vést k vyšší podpoře EU a jejích politik, zatímco negativní rámcování může posilovat euroskeptické postoje a snižovat legitimitu evropských institucí.

3.2 Analýza rámců v kontextu EU

3.2.1 Identifikace hlavních rámců v mediálním pokrytí EU

V kontextu evropské integrace lze identifikovat několik typických rámců, které se objevují v mediálním pokrytí EU. Semetko a Valkenburg (2000) rozlišují pět hlavních rámců:

1. Rámec konfliktu: zdůrazňuje rozpory a střety mezi různými aktéry nebo institucemi
2. Rámec lidského zájmu: klade důraz na osobní příběhy a dopady evropských politik na jednotlivce.
3. Rámec ekonomických důsledků: zaměřuje se na ekonomické aspekty integrace, jako jsou náklady, přínosy nebo dopady na trh práce.
4. Rámec morality: posuzuje evropská témata z hlediska morálních hodnot a principů.
5. Rámec odpovědnosti: přisuzuje odpovědnost za problémy nebo jejich řešení konkrétním aktérům či institucím.

Další autoři identifikovali specifické rámce relevantní pro evropskou integraci, jako je rámec národní suverenity, rámec demokratického deficitu nebo rámec kulturní identity (Galpin a Trenz, 2019). Tyto rámce odrážejí různé aspekty a výzvy evropské integrace, které mohou mít odlišný dopad na veřejné mínění a podporu EU.

Výzkum mediálního pokrytí EU ukazuje, že volba rámců se může lišit mezi jednotlivými médii, zeměmi a časovými obdobími. Jak poznamenávají Boomgaarden et al. (2013), existuje značná variabilita v tom, jak média v různých evropských zemích informují o EU a jaké rámce používají. Tato variabilita může odrážet odlišné národní kontexty, politické kultury a postoje veřejnosti vůči evropské integraci.

3.1.3 Dopad různých rámců na veřejné mínění o evropské integraci

Výzkumy naznačují, že různé rámce mohou mít odlišný dopad na postoje veřejnosti k evropské integraci. Například rámec ekonomických důsledků může vést k vyšší podpoře integrace, pokud jsou zdůrazňovány přínosy jednotného trhu nebo strukturálních fondů (Abbarno a Zapryanova, 2013). Naopak rámec národní suverenity může posilovat euroskeptické postoje, pokud je integrace rámcována jako hrozba pro národní identitu a samostatnost (Leconte, 2015).

Dopad rámcování na postoje veřejnosti k EU však není univerzální a může se lišit mezi jednotlivými zeměmi a skupinami obyvatel. Jak ukazují Gattermann a de Vreese (2020), účinky rámcování mohou být podmíněny celkovým kontextem zpravodajství, důvěryhodností zdrojů nebo individuálními charakteristikami příjemců. Například negativní rámcování EU může mít silnější dopad na postoje jedinců, kteří mají nižší znalosti o evropské integraci nebo jsou k ní skeptičtí.

3.2.3 Dopad různých rámců na veřejné mínění o evropské integraci v České republice

V kontextu České republiky lze očekávat, že dopad různých rámců na veřejné mínění o evropské integraci bude ovlivněn specifickým politickým, ekonomickým a společenským kontextem země. Česká republika se vyznačuje relativně vysokou mírou euroskepticismu ve srovnání s jinými členskými státy EU (Havlík, 2019). Tento euroskepticismus může být posilován rámci, které zdůrazňují negativní aspekty integrace, jako je ztráta národní suverenity, demokratický deficit nebo ekonomické náklady členství v EU.

Zároveň však může být podpora evropské integrace v České republice posilována rámci, které zdůrazňují přínosy členství v EU, jako jsou ekonomické výhody jednotného trhu, možnosti čerpání strukturálních fondů nebo volný pohyb osob (Rakušanová, 2007). Pozitivní rámcování evropské integrace tak může přispívat k vyšší podpoře EU a jejích politik mezi českou veřejností.

Dopad různých rámců na veřejné mínění o evropské integraci v České republice může být ovlivněn také dalšími faktory, jako je politická orientace, socioekonomický status nebo generační příslušnost. Například mladší generace, která vyrostla v době členství České republiky v EU, může být více nakloněna evropské integraci a méně citlivá vůči rámcům

zdůrazňujícím národní suverenitu (Macek et al., 2018). Naopak starší generace, která má zkušenost s obdobím před vstupem do EU, může být více ovlivněna rámci, které zdůrazňují rizika a problémy integrace.

3.3 Koncept sentimentu v mediální analýze

Při analýze mediálního obsahu je důležité rozlišovat mezi sentimentem a postojem (stance), přestože tyto pojmy spolu úzce souvisí. Analýza sentimentu se zaměřuje na identifikaci a kvantifikaci těchto emocí a hodnocení v textu bez ohledu na konkrétní téma nebo cíl. Naproti tomu postoj je definován jako afektivní nebo hodnotící pozice vyjádřená vůči konkrétnímu cíli nebo tématu. V analýze politických textů je obzvláště důležité tyto dva koncepty nezaměňovat, protože politické názory a postoje bývají komplexní a multidimenzionální a mohou být vyjádřeny různými způsoby, pozitivně i negativně (Bestvater a Monroe, 2023).

3.3.1 Definice a význam sentimentu

Sentiment v mediální analýze označuje emocionální zabarvení vyjádřený v mediálním obsahu vůči určitému tématu nebo objektu (Liu, 2015). Analýza sentimentu se zaměřuje na identifikaci a kvantifikaci pozitivních, negativních nebo neutrálních tonů v textu. Sentiment je důležitým ukazatelem toho, jak média prezentují danou problematiku a jaké emoce a hodnocení vůči ní vyjadřují.

Analýza sentimentu vychází z předpokladu, že jazyk není neutrální a že volba slov, frází a stylistických prostředků odráží emoce autora (Taboada, 2016). Sentiment může být vyjádřen explicitně, například pomocí hodnotících přídavných jmen nebo příslovcí (např. "skvělý", "hrozný", "bohužel"), nebo implicitně, prostřednictvím kontextu a konotací (Khoo a Johnkhan, 2018).

Význam analýzy sentimentu spočívá v tom, že poskytuje vhled do emočního rozměru mediálního pokrytí a umožňuje zkoumat, jak média přispívají k utváření veřejného mínění a nálad vůči konkrétním tématům (Taboada, 2016). Převažující sentiment v mediálním pokrytí může ovlivňovat emocionální reakce publika a formovat celkové vyznění veřejné debaty.

3.3.2 Polarita sentimentu a její určování

Polarita sentimentu se obvykle rozlišuje na tři základní kategorie: pozitivní, negativní a neutrální. Pozitivní sentiment vyjadřuje příznivé nebo souhlasné vyjádření vůči tématu, negativní sentiment naopak vyjadřuje kritické nebo odmítavé vyjádření. Neutrální sentiment je charakterizován absencí jasně vyhraněného vyjádření nebo vyváženou prezentací pozitivních a negativních aspektů (Taboada et al., 2011).

Určování polarity sentimentu může probíhat na různých úrovních, od jednotlivých slov a frází až po celé věty a odstavce (Liu, 2015). Na nejnižší úrovni se sentiment určuje na základě tzv. sentimentálních slov, tedy slov, která mají inherentně pozitivní nebo negativní význam (např. "skvělý", "hrozný"). K tomu se využívají předem vytvořené slovníky sentimentálních slov, které přiřazují slovům hodnotu sentimentu (Mohammad, 2016).

Na vyšší úrovni se sentiment určuje na základě kontextu a kombinace různých jazykových prvků, jako jsou negace, intenzifikátory nebo ironické vyjádření (Taboada, 2016). Například věta "Tento film není dobrý" obsahuje negativní sentiment, přestože slovo "dobrý" má samo o sobě pozitivní význam. K zachycení těchto komplexnějších aspektů sentimentu se používají pokročilejší metody, jako jsou strojové učení a hluboké neuronové sítě (Zhang et al., 2018).

Výstupem analýzy polarity sentimentu je obvykle číselné skóre, které vyjadřuje orientaci sentimentu. Například na diskretních proměnných -1 (negativní), 0 (neutrální nebo bez sentimentu) a 1 (pozitivní). Toto skóre umožňuje kvantitativně porovnávat sentiment v různých textech a sledovat jeho vývoj v čase.

Analýza sentimentu však má i své limity a výzvy. Sentiment může být vyjádřen velmi subtilně a ironicky, což může být obtížné zachytit automatickými metodami (Khoo a Johnkhan, 2018). Sentiment také může být subjektivní a závislý na kontextu a interpretaci čtenáře. Proto je důležité kombinovat automatickou analýzu sentimentu s kvalitativním posouzením a zohledňovat širší kontext mediálního pokrytí.

3.4 Vliv sentimentu na veřejné vnímání k EU

Výzkumy naznačují, že sentiment vyjádřený v mediálním pokrytí EU může ovlivňovat veřejné vnímání k evropské integraci. Pozitivní mediální sentiment vůči EU je spojován s

vyšší podporou integrace a důvěrou v evropské instituce, zatímco negativní sentiment může posilovat euroskeptické vnímání (Gattermann a de Vreese, 2020).

Vliv sentimentu na vnímání EU veřejností však není přímočarý a může být podmíněn řadou faktorů. Jak ukazují Brosius et al. (2020), účinky sentimentu mohou záviset na celkovém kontextu zpravodajství, důvěryhodnosti zdrojů nebo individuálních charakteristikách příjemců. Například negativní sentiment může mít silnější dopad na vnímání jedinců, kteří mají nižší znalosti o EU nebo jsou k ní skeptičtí.

Analýza sentimentu v mediálním pokrytí EU také odhaluje zajímavé trendy a rozdíly mezi jednotlivými zeměmi a médii. Jak poznamenávají Galpin a Trenz (2019), sentiment vůči EU se může lišit v závislosti na národním kontextu, politické kultuře a převažujících postojích veřejnosti. Některé země mohou vykazovat konzistentně pozitivnější nebo negativnější sentiment vůči EU než jiné. Podobně mohou existovat rozdíly v sentimentu mezi různými typy médií, například mezi seriózním tiskem a bulvárními deníky.

3.5 Shrnutí kapitoly mediálního rámcování a sentimentu

V kapitole je popsáno využití konceptů, mediálního rámcování a sentimentu, včetně klíčové role, kterou hrají v utváření veřejného mínění o Evropské unii a procesu evropské integrace. Jejich analýza je nezbytná pro pochopení komplexní dynamiky vztahů mezi médii, veřejností a politickými institucemi v kontextu EU. Poznatky z této oblasti mohou přispět k rozvoji teorie evropské integrace a k praktickému využití v oblasti politické komunikace a tvorby politik.

II. Empirická část

Empirická část práce se zaměřuje na aplikaci metod strojového učení pro automatickou analýzu sentimentu a rámců v korpusu českých online mediálních textů o Evropské unii. Cílem je vyvinout a evaluovat modely schopné klasifikovat sentiment a detekovat přítomnost různých rámců v textech, a tím přispět k efektivnější a systematictější analýze mediálního pokrytí evropských témat.

4 Metodologie

4.1 Vytváření korpusu

4.1.1 Výběr zdrojů a období

Vzorek byl vybírán pouze z online médií. Vyhledávání bylo omezeno na 30 českých online médií hodnocených v žebříčku MediaRating. *“MediaRating je nezávislé hodnocení důvěryhodnosti médií. Výsledky hodnocení vyjadřují míru dodržování základních novinářských norem ve vybraných zpravodajských a publicistických médiích v České republice“* (nfnz.cz, 2024). Zahrnuty byly pouze články typu "zpráva" s původem z České republiky. Cílem bylo získat reprezentativní a vyvážený vzorek článků, který by pokrýval různé aspekty evropské integrace a českého členství v EU v období od 1. 1. 2015 do 01.01.2024.

4.1.2 Proces sběru a filtrace dat

Prvním krokem bylo definování klíčových slov pro vyhledávání relevantních článků v archivu Newton Media. Byla použita následující klíčová slova:

1. Názvy institucí a orgánů EU: Evropská unie, EU, Evropská komise, Evropský parlament, Rada Evropské unie, Evropská rada, Soudní dvůr Evropské unie, Evropská centrální banka, Evropský účetní dvůr
2. Vrcholní představitelé EU: předseda Evropské komise, předseda Evropské rady, eurokomisař, europoslanec, předseda Evropského parlamentu
3. Politiky a iniciativy EU: jednotný evropský trh, Schengenský prostor, evropská měnová unie, eurozóna, Erasmus, Horizont Evropa, Zelená dohoda pro Evropu (Green Deal)

4. Obecné termíny: evropská integrace, evropská spolupráce, evropské hodnoty, evropská identita, rozšiřování EU, prohlubování integrace, euroskepticismus, brexit, budoucnost EU, evropské volby, evropské předsednictví, evropské směrnice, evropské nařízení
5. Česko a EU: české předsednictví v Radě EU, Česko a EU, Česká republika a Evropská unie, český eurokomisař, čeští europoslanci, český postoj k EU

Tato klíčová slova byla zadána do vyhledávacího query s logickým operátorem OR.

Následně byla provedena stratifikace článků podle kategorií důvěryhodnosti médií (A, B, C) dle hodnocení MediaRating. Cílem bylo zahrnout stejný počet článků od všech médií prezentovaných v žebříčku MediaRating ve všech kategoriích důvěryhodnosti. V případech, kdy dané médium v určitém roce nepublikovalo články o EU a souvisejících tématech, byl chybějící článek nahrazen článkem z média, které je danému médiu v žebříčku MediaRating nejbližší. Tímto způsobem byl zajištěn vyvážený poměr zastoupení médií z různých kategorií důvěryhodnosti v celkovém korpusu.

Posledním krokem byla manuální kontrola relevance vybraných článků z hlediska jejich tematického zaměření na EU a evropskou integraci. Nerelevantní články byly nahrazeny novými náhodně vybranými články ze stejného média a roku, pokud to bylo možné. V případě nedostatku článků z daného média byly použity články z nejbližší vyšší kategorie důvěryhodnosti (tento případ nastal například u kategorie C, pro médium Britské listy, které byly nahrazeny médii z kategorie B). Všechny vybrané články byly staženy ve formátu XLSX.

4.1.3 Popis výsledného korpusu

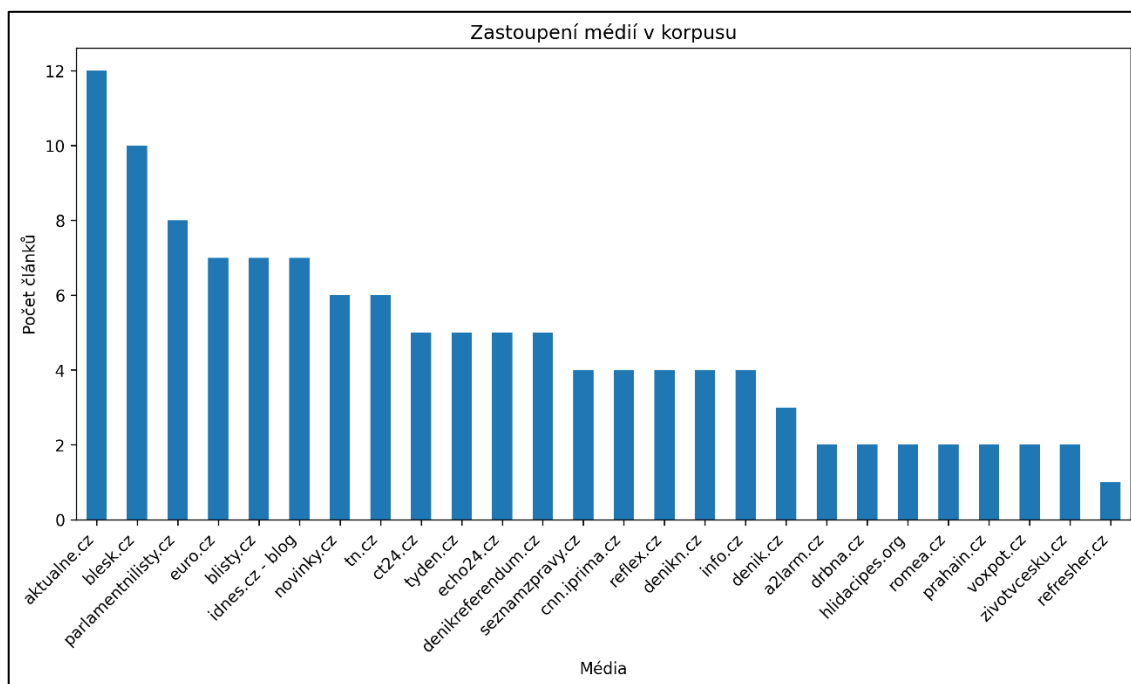
Stažené články byly následně převedeny do jednotného formátu vhodného pro další analýzu. Pomocí skriptu v programovacím jazyce Python byly extrahovány relevantní informace (název článku, text a zdroj) a uloženy do jednotného datového rámce (DataFrame). Výsledný korpus obsahoval celkem 118 článků ze všech zdrojových médií a byl uložen ve formátu CSV.

Výsledný korpus představuje co nejvíce vyvážený a reprezentativní vzorek českého online mediálního pokrytí Evropské unie v období 2015-2024 z 26 médií hodnocených v

žebříčku MediaRating. Korpus pokrývá různá témata související s EU a zohledňuje různé úrovně důvěryhodnosti médií při informování o evropských tématech.

Pro vizualizaci zastoupení jednotlivých médií v korpusu byl vytvořen histogram (viz graf č. 1). Graf poskytuje přehledný obraz o tom, jak jsou různá média v korpusu zastoupena, a umožňuje posoudit vyváženost a reprezentativnost výsledného vzorku.

Graf 1: Zastoupení médií v korpusu



Pečlivá příprava, stratifikace, převedení do jednotného formátu a vizualizace charakteristik korpusu je předpokladem pro validitu, reliabilitu a transparentnost výsledků následné analýzy. Takto vytvořený a popsáný korpus slouží jako základ pro další fáze práce, zejména pro anotaci rámců a sentimentu a pro trénování a testování modelů strojového učení.

4.2 Obsahová analýza

4.2.1 Výběr a popis metody

Pro vytvoření trénovacího datasetu pro analýzu sentimentu a rámců zpravodajských článků o Evropské unii byla v této práci zvolena technika obsahové analýzy. Obsahová analýza je systematická metoda, nebo spíše technika pro objektivní a kvantitativní popis

obsahu textu. Umožňuje identifikovat a kvantifikovat výskyt předem definovaných kategorií v analyzovaném materiálu (Krippendorff, 2004).

Manuální obsahová analýza, při níž výzkumník osobně čte a kóduje text podle stanoveného kódovacího schématu, byla upřednostněna před automatizovanými metodami z několika důvodů. Zaprvé, manuální analýza umožňuje dosáhnout vysoké přesnosti a validity anotací, které jsou klíčové pro trénování kvalitních modelů strojového učení. Zadruhé, manuální kódování dovoluje lépe zachytit kontext a nuance v analyzovaných textech, což je zvláště důležité při identifikaci rámců, které mohou být vyjádřeny implicitně. Automatizované metody oproti tomu spoléhají na předem definovaná klíčová slova a pravidla, což může vést k přehlédnutí důležitých aspektů textu (Krippendorff, 2004).

Hlavním cílem byl stanoven v úvodu této práce. Tento cíl vychází z poznatků odborné literatury, která doporučuje používat metody SML (Supervised machine learning) pro analýzu sentimentu a rámců, neboť při dostatečně velkém datasetu dokáží překonat slovníkové přístupy (Barberá et al., 2021).

Základní soubor pro analýzu tvoří korpus českých online mediálních textů o Evropské unii, jehož sestavení a charakteristiky jsou popsány v kapitole 4.1. Je důležité zmínit, nevyhnutelnost určité míry subjektivity při analýze sentimentu a rámců. Z tohoto důvodu došlo při tvorbě korpusu k jasnému definování kategorií a konzistentní aplikaci kódovacího schématu.

4.2.2 Kódovací schéma

Kódovací schéma použité v této práci se skládá ze dvou hlavních proměnných: polarity sentimentu a přítomnosti rámců. Kódovací jednotkou je segment o délce pěti vět. Tato délka segmentu byla zvolena na základě doporučení Barberá et al. (2021), podle nichž je kódování na úrovni segmentů efektivnější než na úrovni vět, neboť segmenty jsou blíže úrovni objektu, který má být klasifikován. Zároveň segmenty o délce pěti vět poskytují dostatečný kontext pro spolehlivé určení sentimentu a rámců, aniž by byly příliš dlouhé a zahrnovaly více témat. Konkrétní kódovací schéma můžeme vidět v tabulce č. 1.

Třístupňová škála pro kódování polarity sentimentu byla zvolena jako kompromis mezi jednoduchostí a detailností. Zatímco binární klasifikace (pozitivní vs. negativní) by mohla zastříhat důležité nuance v datech, jemnější škála (např. pětistupňová) by mohla snížit

reliabilitu kódování a ztížit trénování klasifikačních modelů (Grimmer a Stewart 2013). Zařazení neutrální kategorie umožňuje odlišit segmenty, které sice zmiňují EU, ale neobsahují zřetelné sentimentální zabarvení.

Tabulka 1: Kódovací schéma

Proměnné	Hodnoty	Názvy hodnot
Polarita sentimentu	-1	negativní
	0	neutrální/ambivalentní
	1	pozitivní
	NA	segment se netýká EU
Rámce	1	Konfliktu
	2	Lidského zájmu
	3	Ekonomických důsledků
	4	Morality
	5	Odpovědnosti
	6	Nelze určit
	NA	Segment se netýká EU

4.2.3 Kódovací instrukce

Kódovací instrukce vyjadřují, jak označovat dané segmenty v textu.

Polarita sentimentu vyjadřuje celkové hodnocení Evropské unie v daném segmentu textu.

- Hodnota -1 značí negativní sentiment vůči EU. Segment je kódován jako negativní, pokud převažují kritické, odmítavé či pesimistické výroky o EU, jejich institucích, politikách nebo představitelích. Příkladem mohou být výroky zpochybňující legitimitu či efektivitu EU, poukazující na její nedostatky a selhání, nebo vyjadřující obavy z negativních dopadů evropské integrace.
- Hodnota 0 označuje neutrální či ambivalentní postoj vůči EU. Segment je kódován jako neutrální, pokud neobsahuje jasně pozitivní ani negativní hodnocení EU, nebo pokud se v něm pozitivní a negativní výroky vyskytují ve vyváženém

poměru. Neutrální segmenty typicky popisují fakta či události související s EU bez explicitního hodnotícího zbarvení.

- Hodnota 1 indikuje pozitivní sentiment vůči EU. Segment je kódován jako pozitivní, pokud převažují příznivé, optimistické či podpůrné výroky o EU, jejich institucích, politikách nebo představitelích. Příkladem mohou být výroky zdůrazňující přínosy a úspěchy evropské integrace vyjadřující důvěru v EU a její budoucnost nebo oceňující konkrétní činy a rozhodnutí unijních orgánů.
- Kód NA (not applicable) se použije v případě, že se segment vůbec netýká Evropské unie. To zahrnuje segmenty pojednávající o čistě vnitrostátních záležitostech bez vztahu k EU nebo o mezinárodních tématech nesouvisejících s evropskou integrací.

Přítomnost rámců je kódována podle typologie Semetko a Valkenburg (2000), která rozlišuje pět obecných rámců často se vyskytujících v politickém zpravodajství, které byli poupraveny k tématu práce, a navíc k nim byli přidány další dva rámce.

1. Rámec konfliktu zdůrazňuje rozpory, neshody a spory mezi jednotlivci, skupinami či institucemi. V kontextu evropské integrace se může jednat například o konflikty mezi členskými státy, evropskými institucemi, politickými frakcemi v Evropském parlamentu či zájmovými skupinami. Segmenty jsou kódovány jako obsahující rámec konfliktu, pokud akcentují rozdílné postoje, zájmy či hodnoty aktérů a vykreslují vztahy mezi nimi jako konfliktní či antagonistické.
2. Rámec lidského zájmu prezentuje témata skrze osobní příběhy a zkušenosti konkrétních jednotlivců. V této práci rámec personalizuje abstraktní procesy a politiky EU tím, že ukazuje jejich dopady na životy obyčejných lidí. Segmenty jsou kódovány jako obsahující rámec zájmu v případě, že se zaměřují na individuální osudy a prožitky ve vztahu k EU, ať už jde o příběhy úspěchu či zklamání.
3. Rámec ekonomický nahlíží na témata optikou ekonomických důsledků a dopadů. V kontextu EU se často zaměřuje na hospodářské přínosy a náklady evropské integrace, fungování jednotného trhu, dopady společných politik či finanční a dluhové krize. Segmenty jsou kódovány jako obsahující ekonomický rámec, pokud primárně hodnotí jednání EU a souvisejících aktérů z hlediska ekonomické výkonnosti, konkurenceschopnosti, růstu či prosperity.
4. Rámec morality posuzuje témata z etického či normativního hlediska. Tento rámec vztahuje jednání EU a evropských aktérů k morálním principům, hodnotám a

ideálům, jako jsou lidská práva, solidarita, spravedlnost či demokratické hodnoty. Segmenty jsou kódovány jako obsahující rámec morality, pokud hodnotí EU a její činy primárně na základě souladu s těmito normativními standardy.

5. Rámec odpovědnosti přisuzuje odpovědnost za řešení problémů a výzev konkrétním aktérům. V kontextu EU se může jednat o odpovědnost evropských institucí, národních vlád, politických stran či jednotlivých politiků. Segmenty jsou kódovány jako obsahující rámec odpovědnosti, pokud jasně identifikují aktéry odpovědné za vznik problémů, jejich řešení, nebo naopak nečinnost tváří v tvář naléhavým výzvám.
6. Kód "Nelze určit" se použije v případě, že segment neobsahuje žádný z výše uvedených rámců. To se může stát například u ryze deskriptivních segmentů bez interpretačního rámce, nebo u segmentů využívajících velmi specifické, idiosynkratické rámce.
7. NA (not applicable) se použije ve stejném případě jako u aplikace hodnot sentimentu. To znamená, že není v segmentu patrný žádný kontext na EU.

Kategorie rámců jsou pojaty jako konjunkční, tj. každému segmentu může být přiřazen žádný, jeden nebo více rámců. Toto rozhodnutí vychází z poznatku, že jednotlivé texty často obsahují vícero vzájemně se překrývajících a doplňujících rámců (Matthes a Kohring, 2008).

4.2.4 Proces a výsledek anotace dat

Anotační proces dat probíhal pomocí vlastního skriptu vytvořeného v programovacím jazyce Python. Skript načítal jednotlivé články z korpusu a rozděloval je na segmenty o délce pěti vět, které sloužily jako kódovací jednotky. Tyto segmenty byly postupně prezentovány anotátorovi v konzoli. Anotování článků prováděl jen jeden anotátor.

Anotátor měl za úkol každému zobrazenému segmentu přiřadit hodnoty dvou proměnných: polaritu sentimentu a přítomnosti rámců, v souladu s kódovacím schématem popsaným v kapitole 4.2.2 a 4.2.3. Anotátor vybíral příslušné hodnoty zadáním odpovídajících kódů. Skript průběžně ukládal anotované segmenty a jejich příslušnost k původním článkům.

Výsledný klasifikovaný korpus vypadá následovně. Z celkových 118 článků bylo vytvořeno 907 segmentů, které sice mají jiný počet slov, ale splňují kritérium pěti vět na jeden segment. Rozložení sentimentu je prezentováno v tabulce č. 2. Nejčastěji se vyskytoval

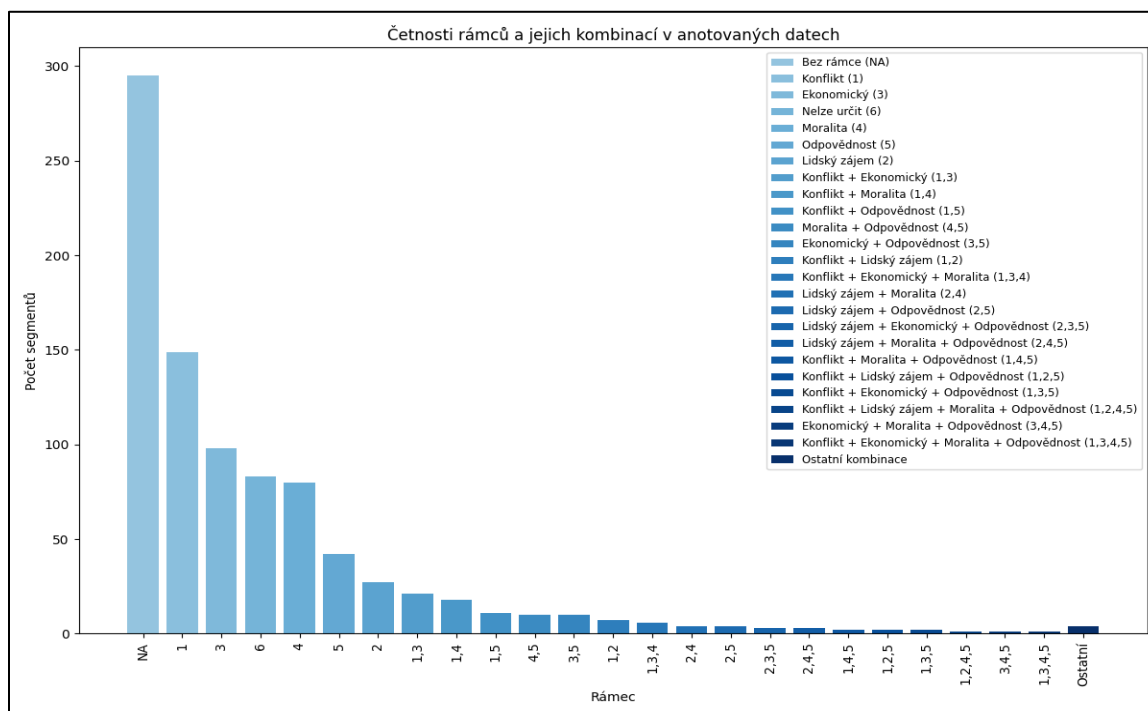
neutrální sentiment, který byl následován segmenty, jenž se netýkali EU. V tabulce číslo 2 můžeme vidět že 154 segmentů bylo označeno jako negativní sentiment, zatímco 34 segmentů mělo pozitivní sentiment. Nejvyšší četnost byla naměřena u neutrálního sentimentu, který odpovídal 424 segmentům.

Tabulka 2: Rozložení sentimentu v anotovaných datech

Polarita Sentimentu	Počet segmentů
Neutrální (0)	424
Netýká se EU (NA)	295
Negativní (-1)	154
Pozitivní (1)	34

Graf 2 ukazuje četnosti výskytu jednotlivých rámců a jejich kombinací v anotovaných datech. Nejčastěji se vyskytovaly segmenty, které se netýkali EU (295 segmentů), následované rámcem konfliktu (149 segmentů) a ekonomickým rámcem (98 segmentů). Ostatní rámce a jejich kombinace se vyskytovaly s nižší frekvencí. Zajímavé je, že se v datech objevily i segmenty s třemi či čtyřmi souběžnými rámci, byť v malém počtu.

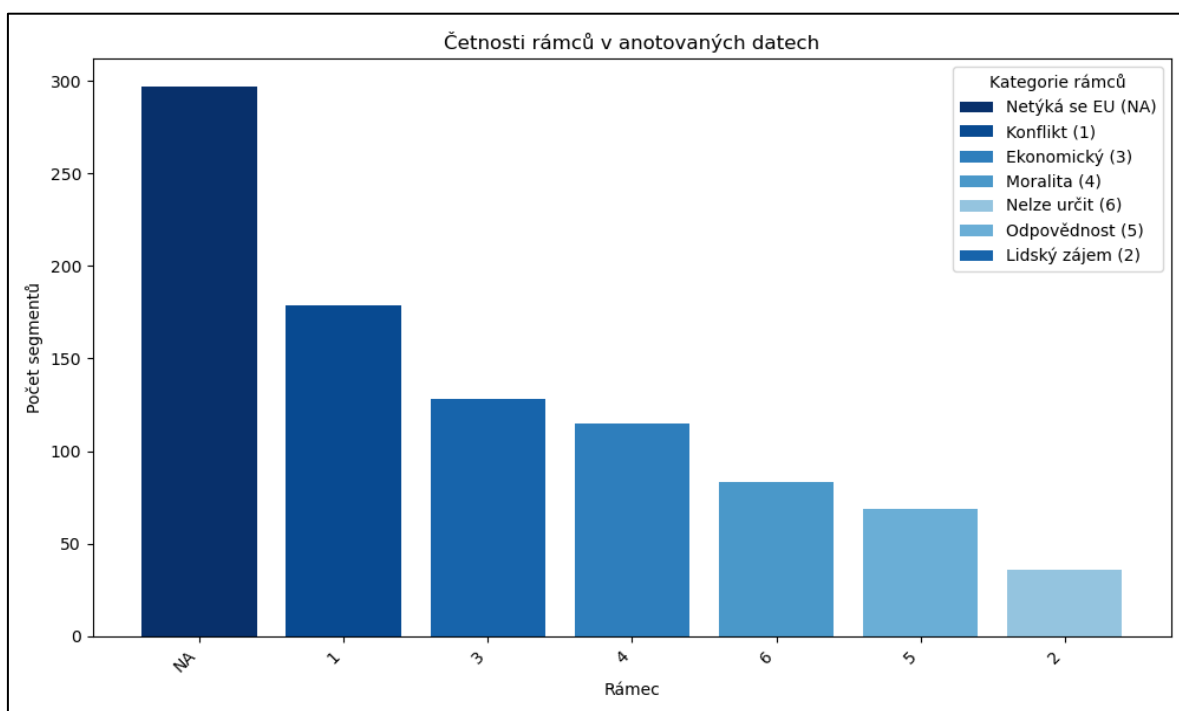
Graf 2: Četnost rámců v korpusu



Převaha neutrálního sentimentu a segmentů které se netýkali EU naznačuje, že značná část korpusu je spíše deskriptivní a faktická. Zároveň je patrná polarizace mezi negativním a pozitivním sentimentem, s výraznou převahou negativního. Výskyt kombinací rámců, byť v menší míře, poukazuje na komplexnost a mnoho vrstevnatost mediálních reprezentací EU. Častý výskyt rámce konfliktu pak může odrážet konflikty a spory mezi jednotlivými aktéry o směřování evropské integrace a rolích jednotlivých aktérů.

Pro účely další analýzy bylo rozhodnuto použít pouze jeden rámec na segment článku s cílem zjednodušit analýzu a zajistit konzistenci dat napříč všemi segmenty. V případech, kdy byl segment původně označen více rámci (multilabel), byla provedena redukce na jediný rámec. Pro segmenty s více rámci byl aplikován algoritmus náhodného výběru, kdy byl z přiřazených rámců náhodně vybrán jeden, který byl ponechán jako jediný reprezentant daného segmentu, zatímco ostatní rámce byly z anotace odstraněny. Tento přístup zajistil, že každý segment je nyní reprezentován právě jedním rámcem, což umožňuje jednodušší a přímočařejší analýzu dat, avšak je třeba brát v úvahu potenciální ztrátu některých informací v původních datech. V grafu č. 3 je vidět finální rozložení rámců pro další analýzu.

Graf 3: Revidované četnosti rámců v korpusu



V tabulce č.3 jsou uvedeny podrobné četnosti jednotlivých kategorií rámců, které byly použity pro další analýzu. Nejčastěji zastoupenou kategorií je "Netýká se EU (NA)" s 297

segmenty, což naznačuje, že část segmentů korpusu se primárně nezabývá tématy spojenými s Evropskou unií. Druhým nejčastějším rámcem je "Konflikt (1)" se 179 segmenty, následovaný "Ekonomickým (3)" rámcem se 128 segmenty. Tyto výsledky poukazují na to, že v korpusu jsou často diskutovány konflikty a ekonomické otázky související s EU. Rámce "Moralita (4)" a "Nelze určit (6)" jsou zastoupeny 115 a 83 segmenty. Nejméně častým rámcem je "Lidský zájem (2)" s 36 segmenty.

Tabulka 3: Četnosti rámců v korpusu

Kategorie rámců	Počet segmentů
Netýká se EU (NA)	297
Konflikt (1)	179
Ekonomický (3)	128
Moralita (4)	115
Nelze určit (6)	83
Odpovědnost (5)	69
Lidský zájem (2)	36

5 Strojové učení

5.1 Předzpracování textu

V této práci jsou použity dva přístupy pro reprezentaci textových dat: SBERT (Sentence-BERT) embeddings a Bag of Words (BoW) model. Každý z těchto přístupů vyžaduje odlišné kroky předzpracování textu.

Pro transformaci textových dat na numerické vektory pomocí modelu SBERT není nutné provádět rozsáhlé předzpracování textu. SBERT byl trénován na velkých neupravených textových korpusech a je schopen si sám odvodit sémanticky relevantní informace z původních tvarů slov a jejich kontextu. Samotná transformace textových dat na numerické vektory pomocí SBERT bude popsána v kapitole 5.2.

Na rozdíl od SBERT embeddings vyžaduje Bag of Words model více kroků předzpracování textu. Předzpracování textu pro BoW model se skládá z následujících kroků:

1. Tokenizace: Text je rozdělen na jednotlivá slova (tokeny) pomocí tokenizátoru. Tokenizátor bere v úvahu specifika daného jazyka a odděluje slova od interpunkčních znamének.
2. Odstranění stop slov: Ze seznamu tokenů jsou odstraněna stop slova, což jsou častá slova, která nenesou významovou informaci (např. předložky, spojky). Odstranění stop slov pomáhá snížit komplexitu BoW reprezentace a zaměřit se na důležitá slova.
3. Lemmatizace: Slova jsou převedena na jejich základní tvar (lemma) pomocí lemmatizátoru. Lemmatizace sjednocuje různé tvary stejného slova (např. "běžel", "běží", "poběží" na "běžet"), což zjednodušuje reprezentaci textu a snižuje počet příznaků.

Výstupem fáze předzpracování textu jsou dva typy korpusů: Jeden pro SBERT embeddings a druhý pro BoW matice příznaků. Tyto reprezentace jsou dále použity pro vytvoření specifických vektorů, které poslouží k trénování a evaluaci modelů strojového učení pro klasifikaci sentimentu a detekci rámců.

5.2 Transformace textu na numerické vektory pomocí SBERT

Pro transformaci předzpracovaných textových segmentů na numerické vektory je použit model SBERT (Sentence-BERT). SBERT je modifikovaná verze architektury BERT

(Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019), která je optimalizovaná pro vytváření vektorových reprezentací celých vět či segmentů textu.

Hlavní výhodou SBERT oproti původnímu BERT je, že dokáže generovat embeddings pro celé segmenty přímo, bez nutnosti dodatečného zprůměrování embeddings jednotlivých slov. SBERT je trénován pomocí dvojic vět s cílem maximalizovat kosinovou podobnost embeddings pro sémanticky podobné věty a minimalizovat ji pro nepodobné věty. Tento přístup vede k vytvoření husté vektorové reprezentace, která lépe zachycuje sémantický obsah a kontext segmentů (Reimers a Gurevych, 2019). Zvolený model SBERT s přesným názvem "paraphrase-xlm-r-multilingual-v1" byl trénován na vícejazyčných datech pomocí architektury XLM-RoBERTa zahrnující i český jazyk (Conneau et al., 2020). V této práci využíváme Google Colab pro implementaci daného typu SBERT. Google Colab je cloudovou platformou pro strojové učení, která nabízí přístup k výkonnému hardwaru a usnadňuje práci v jejím prostředí.

Transformace segmentů na numerické vektory probíhá následovně v teoretické rovině:

1. Každý předzpracovaný segment (v českém jazyce) je předán modelu SBERT jako vstup.
2. Model vytvoří embeddings vektor pro celý segment, který zachycuje jeho sémantický obsah a kontext. Tento vektor má fixní délku 768 dimenzí bez ohledu na délku vstupního segmentu.
3. Výsledné embeddings jsou normalizovány na jednotkovou délku (L2 normalizace). Normalizace zlepšuje stabilitu a konvergenci modelů strojového učení a umožňuje měřit podobnost segmentů pomocí kosinové vzdálenosti (úhel mezi vektory).

Výstupem této fáze je datová sada, kde je každý segment reprezentován numerickým vektorem o 768 dimenzích s normalizovanými hodnotami. Segmenty jsou opatřeny příslušnými anotacemi sentimentu a rámců z předchozí fáze obsahové analýzy. Tato datová sada je nyní připravena pro trénování a evaluaci modelů strojového učení pro klasifikaci sentimentu a detekci rámců. Proces zpracování embeddings v google colab je podrobně popsán v příloze práce.

5.3 Transformace textu na numerické vektory pomocí BoW

Bag of Words (BoW) je jednoduchá, ale účinná technika pro reprezentaci textových dat v úlohách zpracování přirozeného jazyka. V modelu BoW je text reprezentován jako neuspořádaná množina slov, přičemž se nebere v úvahu pořadí slov ani gramatická struktura. Každý dokument je převeden na vektor příznaků, kde každý prvek vektoru představuje počet výskytů daného slova v dokumentu.

V této práci používáme strategii BoW pro transformaci předzpracovaných textových segmentů na numerické vektory. K vytvoření BoW reprezentace používáme třídu `CountVectorizer` z knihovny `scikit-learn`. Proces vytvoření BoW matice se skládá z následujících kroků:

1. Načtení předzpracovaných dat z CSV souboru, který obsahuje lemmatizované segmenty textů, názvy článků, sentiment a rámce.
2. Extrakce lemmatizovaných segmentů a jejich spojení do jednotlivých textových řetězců. Současně se replikují odpovídající názvy článků, sentiment a rámce pro každý segment.
3. Vytvoření instance třídy `CountVectorizer` a provedení transformace textových řetězců segmentů na BoW matici. Každý řádek matice představuje jeden segment a každý sloupec reprezentuje unikátní slovo (příznak) v korpusu.
4. Převod BoW matice na hustou numpy matici a následně na `DataFrame`, kde sloupce představují jednotlivá slova (příznaky) a řádky reprezentují segmenty.
5. Přidání sloupců `'Title'`, `'Sentiment_Polarity'` a `'Frames'` do `DataFrame`, které obsahují odpovídající informace pro každý segment.
6. Uložení výsledného `DataFrame` s BoW reprezentací do nového CSV souboru.

Výsledkem tohoto procesu je datová sada, kde je každý segment reprezentován numerickým vektorem příznaků, který zachycuje počet výskytů jednotlivých slov v daném segmentu.

Kompletní implementace procesu vytvoření BoW matice v jazyce Python je uvedena v příloze práce.

5.4 Výběr a popis modelu strojového učení

Pro úlohu automatické analýzy sentimentu a rámců byl zvolen model strojového učení Support Vector Machines (SVM). Tento model byl vybrán pro svou dobrou výkonnost a interpretovatelnost v úlohách klasifikace textů.

SVM je diskriminativní klasifikátor, který hledá nadrovinu v příznakovém prostoru maximalizující vzdálenost mezi třídami. Klíčovým konceptem SVM jsou jádrové funkce (kernel functions), které umožňují efektivně převést původní příznakový prostor do vyšší dimenze, kde jsou třídy lineárně separovatelné. V této práci je použit RBF (Radial Basis Function) kernel, který se osvědčil jako zvláště účinný v situacích, kde počet dimenzí převyšuje počet vzorků, což je běžné v textových úlohách. V takových případech jsou příznaky často reprezentovány pomocí slovních vektorů vysoké dimenzionality. RBF jádro umožňuje modelům efektivně pracovat v nelineárních prostorech a je schopné mapovat původní prostor příznaků do vyšších, dokonce nekonečných dimenzí. Tento proces zjednodušuje oddělení tříd prostřednictvím hyperroviny, což je zásadní pro účinnou klasifikaci v komplexních datových souborech (Awad a Khanna, 2015).

SVM má ve srovnání s ostatními běžně používanými klasifikačními nástroji pro strojové učení, jako jsou rozhodovací stromy, k-nejbližších sousedů (k-NN) nebo Naive Bayes, několik výhod i nevýhod. Jednou z hlavních výhod SVM je jeho vysoká přesnost, zejména v úlohách s vysokou dimenzionalitou příznaků, jako je klasifikace textů (Pawar a Gawande, 2012). Díky principu maximalizace okraje mezi třídami je SVM také méně náchylný k přeučení než některé jiné modely. Použití jádrových funkcí umožňuje SVM efektivně zachytit nelineární vztahy mezi příznaky bez nutnosti explicitní transformace příznakového prostoru. Navíc SVM poskytuje informace o důležitosti jednotlivých příznaků pro klasifikaci, což může přispět k interpretovatelnosti modelu (Hsieh a Dhillon, 2014).

Na druhou stranu, trénování SVM může být výpočetně náročné, zejména pro velké datové sady s mnoha příznaky (Hsieh a Dhillon, 2014). Výkonnost SVM také silně závisí na volbě hyperparametrů, jako je typ jádrové funkce a regularizační parametr C . Nalezení optimálních hodnot může vyžadovat rozsáhlé ladění. Interpretace modelu SVM může být náročná v případě velmi vysokého počtu příznaků. Kromě toho SVM předpokládá, že jsou třídy lineárně separovatelné v transformovaném příznakovém prostoru, což nemusí být vždy splněno v reálných datech (Zulfiqar et al., 2022).

Při výběru klasifikačního modelu je proto důležité zvážit charakteristiky konkrétní úlohy a datové sady a porovnat výkonnost SVM s ostatními modely pomocí křížové validace nebo jiných technik evaluace (Conneau et al., 2020). V kontextu této práce se SVM jeví jako vhodná volba vzhledem k vysoké dimenzionalitě příznaků získaných pomocí SBERT embeddings a relativně malému počtu trénovacích vzorků.

5.5 Implementace SVM modelu s SBERT transformací

Modely SVM jsou trénovány na 70 % dat z výchozí datové sady, která byla v předchozích krocích anotována a transformována na numerické vektory pomocí SBERT embeddings. Zbývajících 30 % dat slouží jako testovací množina pro evaluaci modelu. Rozdělení dat na trénovací a testovací množinu je provedeno pomocí funkce `train_test_split` z knihovny `scikit-learn`.

Před samotným trénováním je provedena úprava cílových proměnných (sentiment a rámce) tak, aby byly kompatibilní s požadavky modelu SVM. U sentimentu jsou hodnoty 'NA' nahrazeny číselnou hodnotou 99 a u rámců hodnotou 0. Následně je pro každý článek vybrán nejčastěji se vyskytující sentiment a rámec pomocí funkce `mode` z modulu `statistics`. Tímto způsobem získáme jednu hodnotu sentimentu a rámce pro každý článek, což umožňuje trénování modelů na úrovni článků. Tato úprava byla provedena vzhledem k náročnosti implementace SBERT embeddings pro každý segment. Toto rozhodnutí je dvojsečné, jelikož sice získáváme přehled o rozdělení rámců a sentimentu k určitým článkům. Na druhou stranu zde dochází ke ztrátě informace, která by mohla být potenciálně prospěšná k lepší výkonnosti modelu.

Modely jsou trénovány odděleně pro úlohu klasifikace sentimentu a pro úlohu detekce rámců. V obou případech je použit model SVM s jádrem RBF (Radial Basis Function) a parametrem `class_weight` nastaveným na "balanced", který automaticky upravuje váhy tříd nepřímo úměrně jejich četnostem. Tento přístup pomáhá při práci s nevyváženými daty, jako jsou ta v datové sadě, které používáme v práci.

Ladění hyperparametrů modelů není provedeno pomocí grid search a křížové validace. Z důvodu výpočetní náročnosti a nelehké implementace do kódu. Místo toho jsou modely trénovány s výchozími hodnotami hyperparametrů.

5.6 Evaluace výkonnosti modelu SVM pro SBERT

Po natrénování modelů je jejich výkonnost vyhodnocena na testovací množině pomocí standardních metrik pro klasifikační úlohy: přesnost (accuracy), preciznost (precision), úplnost (recall) a F1 skóre. V případě klasifikace sentimentu dosáhl model SVM následujících výsledků. Tabulka č. 4.

Tabulka 4: Výsledky SVM modelu pro klasifikaci polarity sentimentu s SBERT transformací

Metriky	Výsledky
Přesnost	0.703
Preciznost	0.741
Úplnost	0.703
F1 skóre	0.660

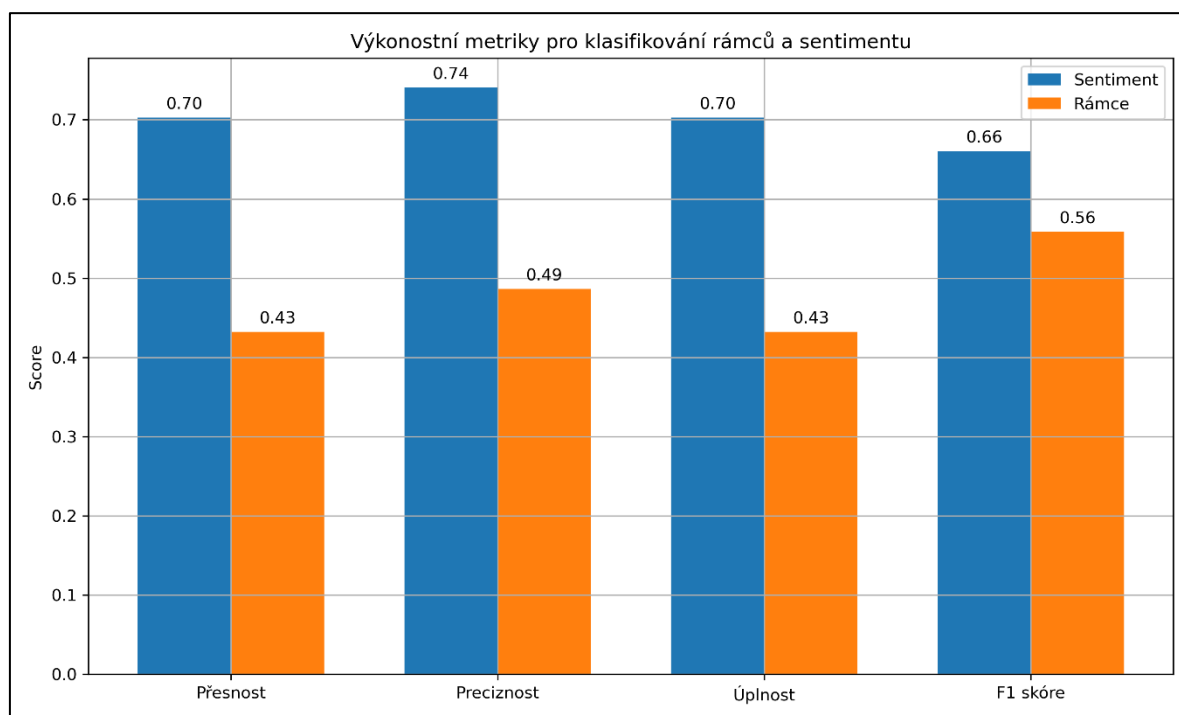
Tyto výsledky naznačují, že model pro klasifikaci sentimentu dosahuje relativně dobré výkonnosti, s přesností a úplností přes 70 % a F1 skóre přes 66 %. Model je schopen správně rozlišovat mezi různými třídami sentimentu a má vyváženou výkonnost z hlediska preciznosti a úplnosti. U úlohy detekce rámců byly dosaženy následující výsledky, které jsou uvedeny v Tabulce č. 5.

Tabulka 5: Výsledky SVM modelu pro klasifikaci rámců s SBERT transformací

Metriky	Výsledky
Přesnost	0.432
Preciznost	0.486
Úplnost	0.432
F1 skóre	0.558

Výsledky pro detekci rámců jsou méně uspokojivé než pro klasifikaci sentimentu, s přesností a úplností kolem 43 % a F1 skóre kolem 56 %. To naznačuje, že úloha detekce rámců je náročnější a vyžaduje další vylepšení modelu a přístupu. V grafu č. 4 je vidět srovnání metrik pro dané klasifikační úlohy.

Graf 4: Porovnání metrik pro klasifikaci polarity sentimentu a klasifikaci rámců



5.7 Implementace SVM modelu s BoW transformací

Implementace modelu využívajícího Bag of Words (BoW) reprezentaci byla provedena s cílem porovnat jeho výkonnost s modelem založeným na SBERT embeddings. Pro zajištění férovosti srovnání jsme použili stejný dataset a obdobný metodologický přístup jako v případě SBERT. Proces trénování a ladění modelu Support Vector Machines (SVM) pro BoW reprezentaci zahrnoval několik klíčových kroků, které byly navrženy tak, aby co nejvíce odpovídaly postupu použitému u SBERT embeddings.

Prvním krokem byla příprava dat, která začala načtením předzpracovaných textových dat z CSV souboru. Tento soubor obsahoval BoW reprezentace jednotlivých segmentů textu, názvy článků, sentimentu a rámce. Aby se zachovala kontinuita a férovost pro srovnání modelů, bylo nutné BoW data také agregovat na úroveň článků. To jsme provedli seskupením dat podle názvu článku, přičemž jsme pro každý článek vybrali nejčastější sentiment a rámec. BoW vektory všech segmentů v rámci jednoho článku byly sečteny, čímž jsme získali reprezentaci celého článku.

Následně jsme přistoupili k úpravě cílových proměnných. Hodnoty Na v sloupcích sentimentu a rámců byly nahrazeny hodnotami 99 pro sentiment a 0 pro rámce, což

odpovídalo přístupu použitému u SBERT embeddings. Tato úprava zajistila konzistenci dat a umožnila modelu pracovat se všemi proměnnými datasetu.

Data byla poté rozdělena na trénovací a testovací množinu v poměru 70:30 pomocí funkce `train_test_split` z knihovny `scikit-learn`. Tento krok zajistil, že model bude hodnocen na dosud neviděných datech, což umožňuje realistické posouzení jeho generalizační schopnosti. Zachování stejného poměru rozdělení jako u SBERT přístupu dále přispělo k porovnatelnosti obou metod.

Pro samotné trénování jsme zvolili model SVM s lineárním jádrem, a to jak pro klasifikaci sentimentu, tak pro detekci rámců. Volba lineárního jádra byla motivována vysokou dimenzionalitou BoW reprezentace a relativně malým počtem trénovacích vzorků. Důležitým aspektem nastavení modelu bylo použití parametru `class_weight="balanced"`, který automaticky upravuje váhy tříd nepřímo úměrně jejich četnostem v trénovacích datech. Toto nastavení pomáhá modelu lépe se vypořádat s případnou nevyvážeností tříd v datasetu.

Evaluace natrénovaných modelů byla provedena na testovací množině s využitím standardních metrik pro klasifikační úlohy: přesnost (accuracy), preciznost (precision), úplnost (recall) a F1 skóre.

5.8 Evaluace výkonnosti SVM s BoW

Po natrénování modelů jsme provedli evaluaci jejich výkonnosti na testovací množině. Výsledky jsou vidět v tabulce č. 6.

Tabulka 6: Výsledky SVM modelu pro klasifikaci polarity sentimentu s BoW transformací

Metriky	Výsledky
Přesnost	0.722
Preciznost	0.705
Úplnost	0.722
F1 skóre	0.681

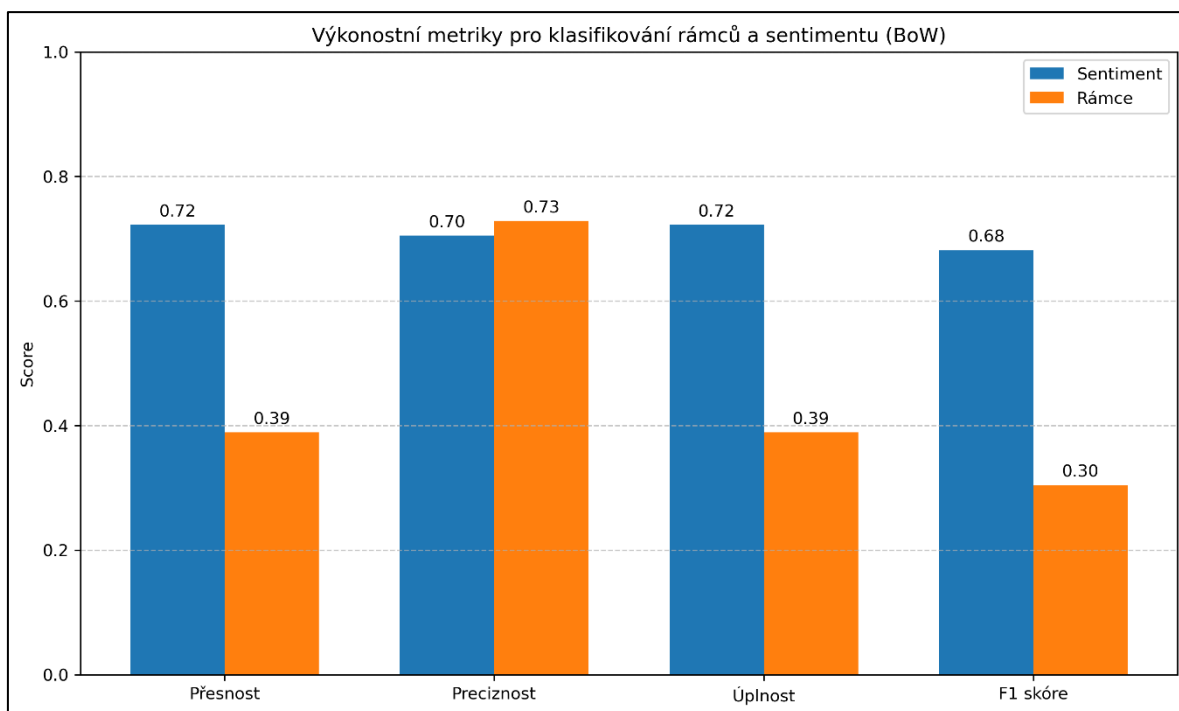
Tyto výsledky ukazují poměrně dobrou výkonnost modelu pro klasifikaci sentimentu. Vyrovnané hodnoty přesnosti a úplnosti naznačují, že model je konzistentní ve své schopnosti správně identifikovat a klasifikovat sentiment napříč různými třídami.

Tabulka 7: Výsledky SVM modelu pro klasifikaci rámců s BoW transformací

Metriky	Výsledky
Přesnost	0.389
Preciznost	0.728
Úplnost	0.389
F1 skóre	0.304

Výsledky pro klasifikaci rámců nejsou tak příznivé jako u klasifikace sentimentu. Nízká přesnost a úplnost v kombinaci s vysokou precizností naznačují, že model je velmi selektivní při klasifikaci rámců. V grafu č. 5 je vidět vizuální srovnání metrik pro rámce a sentiment s použitím BoW transformace.

Graf 5: Přehled výkonostních metrik pro SVM model s BoW transformací



6 Srovnání výsledků SBERT a BoW transformací

Analýza výsledků modelů využívajících SBERT embeddings a Bag of Words (BoW) reprezentaci poskytuje cenné poznatky o efektivitě těchto metod při zpracování mediálního obsahu o Evropské unii v českém kontextu.

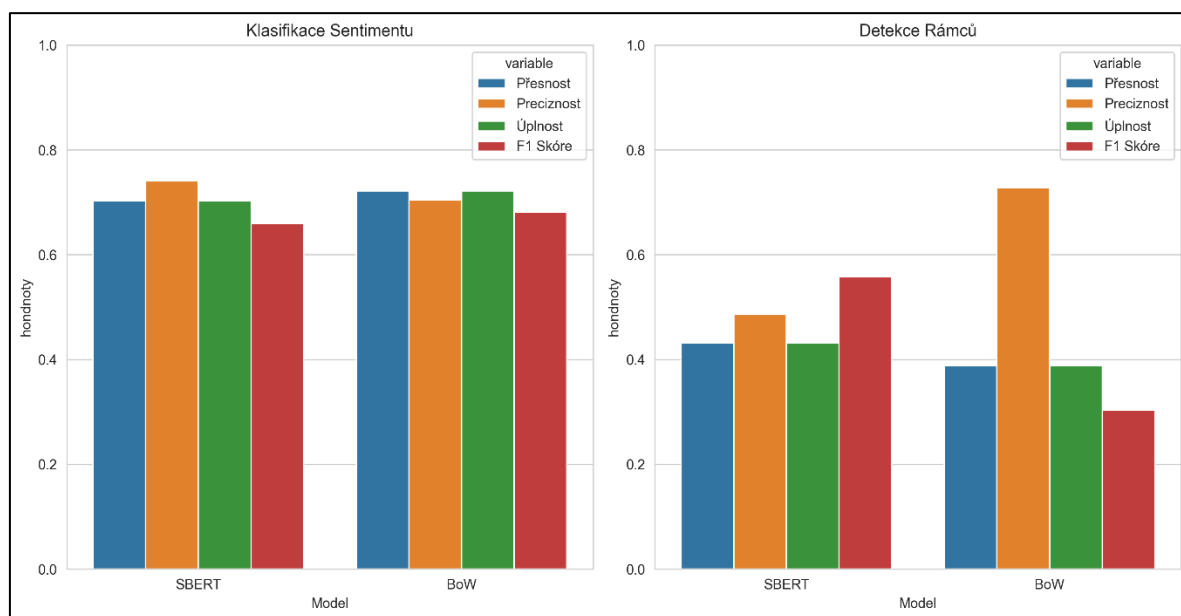
Klasifikace sentimentu: SBERT model dosáhl přesnosti 70,3 %, zatímco model s použitím BoW mírně překonal tento výsledek s přesností 72,2 %. Tento rozdíl, ačkoli malý, je zajímavý vzhledem k rozložení sentimentu v našem korpusu. S ohledem na to, že velký počet článků bylo označeno jako netýkající se EU (58 %), oba modely prokázaly schopnost efektivně rozlišovat mezi touto převládající kategorií a méně zastoupenými pozitivními negativními a neutrálními sentimenty.

Mírně lepší výkon BoW modelu v této úloze může být vysvětlen tím, že pro klasifikaci sentimentu mohou být klíčová specifická slova nebo fráze, které BoW reprezentace dokáže efektivně zachytit. SBERT, ačkoli sofistikovanější, může v některých případech přehlédnout tyto specifické indikátory ve snaze zachytit celkový kontext.

V úloze detekce rámců SBERT model předčil BoW model, dosáhnuv přesnosti 43,2 % oproti 38,9 % u BoW. Tento rozdíl je významný a odráží komplexnost úlohy detekce rámců, která vyžaduje hlubší porozumění kontextu a sémantice textu.

Rozložení rámců v našem korpusu bylo poměrně nevyvážené, s většinovým zastoupením článků netýkajících se EU. Schopnost SBERT modelu dosáhnout vyšší přesnosti v této náročné úloze ukazuje na jeho lepší schopnost zachytit jemné nuance a kontextové informace potřebné pro správnou identifikaci rámců. Přehledné srovnání všech výsledků je vidět v grafu č. 6.

Graf 6: Grafické srovnání metrik, při použití různých transformací textu



Oba modely prokázaly solidní schopnost klasifikovat sentiment. Mírně lepší výkon BoW modelu naznačuje, že pro tuto konkrétní úlohu může být jednodušší lexikální přístup dostačující a v některých případech vhodnější.

Výrazně lepší výkon SBERT modelu v detekci rámců poukazuje, že sofistikovanější metody jsou vhodnější pro zachycení komplexních konceptuálních struktur, jako jsou mediální rámce. Tento výsledek je obzvláště cenný vzhledem k nevyváženosti rámců v našem korpusu.

Tato práce demonstruje potenciál metod strojového učení pro automatickou analýzu mediálního obsahu o EU. Vyvinuté modely, zejména SBERT model pro detekci rámců, představují významný nástroj v systematické a efektivní analýze velkých objemů mediálních dat. Dosažené výsledky korespondují s našimi původními cíli. Podařilo se nám vyvinout robustní klasifikační modely, které dokáží spolehlivě identifikovat sentiment a tematické rámce v českých mediálních textech o EU. Zejména v případě detekce rámců náš SBERT model skutečně předčil tradiční metodu BoW, což bylo jedním z hlavních záměrů této práce. Naše zjištění přispívají k lepšímu pochopení možností a limitů různých přístupů k analýze mediálního obsahu v kontextu evropské integrace. Výsledky ukazují, že zatímco pro jednodušší úlohy jako klasifikace sentimentu mohou tradiční metody stále poskytovat konkurenceschopné výsledky, pro komplexnější úlohy jako je identifikace rámců jsou sofistikovanější metody jako SBERT vhodnější.

Závěr

Tato bakalářská práce se zabývala analýzou mediálního obrazu Evropské unie v českém kontextu s využitím metod strojového učení. Hlavním cílem bylo vyvinout robustní klasifikační model schopný identifikovat sentiment a tematické rámce v mediálních textech týkajících se EU a porovnat jeho výkonnost s tradičním přístupem založeným na metodě Bow.

Teoretická část práce poskytla konceptuální rámec pro analýzu mediálních reprezentací EU. Byla diskutována role médií v informování veřejnosti o evropských tématech a jejich vliv na vnímání EU ze strany občanů. Představeny byly koncepty mediálního rámcování a sentimentu jako klíčové nástroje pro analýzu mediálního obsahu a prozkoumány možnosti jejich zkoumání pomocí metod strojového učení s učitelem. Tato část zdůraznila význam zkoumání mediálních reprezentací EU pro pochopení dynamiky veřejného mínění a politických procesů souvisejících s evropskou integrací.

V empirické části byl popsán proces vytvoření korpusu českých online mediálních textů o EU, zahrnující výběr zdrojů, sběr a filtraci dat a anotaci sentimentu a rámců pomocí manuální obsahové analýzy. Anotovaná data byla následně předzpracována a transformována na numerické vektory pomocí modelu SBERT pro trénování a evaluaci klasifikačních modelů.

Hlavním přínosem této práce je vytvoření funkčního modelu pro automatickou analýzu sentimentu a rámců v českých mediálních textech o EU. Tento model může být využit pro efektivnější a systematictější monitorování mediálního pokrytí evropských témat a poskytovat cenné poznatky o dynamice veřejného mínění vůči EU. Práce také přispívá k rozvoji metodologie analýzy mediálního obsahu pomocí technik strojového učení a demonstruje potenciál těchto přístupů pro zkoumání komplexních společenských a politických jevů.

Je však třeba zmínit i limity této práce. Jedním z omezení je relativně malá velikost datového souboru použitého pro trénování a evaluaci modelů. Větší a rozmanitější datový soubor by mohl vést k vyšší přesnosti a robustnosti modelů. Dalším limitem je značné redukování informací vytvořeného korpusu pro aplikaci SML modelu, který by bez těchto modifikací byl těžko aplikovatelný. Výsledky mohou být také ovlivněny subjektivitou anotátora při manuální obsahové analýze, ačkoli byla snaha o maximální konzistenci a reliabilitu anotací.

Budoucí výzkum by se mohl zaměřit na další vylepšení výkonnosti modelu, například využitím pokročilejších architektur neuronových sítí nebo zahrnutím dodatečných lingvistických příznaků. Nebo také o vytvoření sady instrukcí, který by umožnil využít potenciál všech anotovaných segmentech. Zajímavé by také bylo rozšířit analýzu na delší časové období a větší vzorek médií, aby bylo možné sledovat vývoj mediálního obrazu EU v čase a identifikovat případné trendy a zlomy. Další oblastí pro budoucí výzkum je integrace analýzy mediálního obsahu s dalšími typy dat, jako jsou průzkumy veřejného mínění nebo data ze sociálních médií, s cílem získat komplexnější pohled na vztah mezi mediálními reprezentacemi, veřejným míněním a politickým rozhodováním v kontextu evropské integrace.

Seznam zdrojů

ABBARNO, Aaron J. a Galina M. ZAPRYANOVA, 2013. Indirect Effects of Eurosceptic Messages on Citizen Attitudes toward Domestic Politics. *JCMS: Journal of Common Market Studies* [online]. **51**(4), 581–597 [vid. 2024-04-26]. ISSN 1468-5965. Dostupné z: doi:10.1111/jcms.12025

ANDERSEN, Kim, Claes H. DE VREESE a Erik ALBÆK, 2016. Measuring Media Diet in a High-Choice Environment - Testing the List-Frequency Technique. *Communication Methods and Measures* [online]. **10**(2–3), 81–98 [vid. 2024-04-25]. ISSN 1931-2458. Dostupné z: doi:10.1080/19312458.2016.1150973

ANDERSON, Christopher J., 1998. When in Doubt, Use Proxies: Attitudes toward Domestic Politics and Support for European Integration. *Comparative Political Studies* [online]. **31**(5), 569–601 [vid. 2024-04-25]. ISSN 0010-4140. Dostupné z: doi:10.1177/0010414098031005002

ANDERSON, CHRISTOPHER J. a KARL C. KALTENTHALER, 1996. The Dynamics of Public Opinion toward European Integration, 1973-93. *European Journal of International Relations* [online]. **2**(2), 175–199 [vid. 2024-04-25]. ISSN 1354-0661. Dostupné z: doi:10.1177/1354066196002002002

ANDERSON, Christopher J. a M. Shawn REICHERT, 1995. Economic Benefits and Support for Membership in the E.U.: A Cross-National Analysis. *Journal of Public Policy* [online]. **15**(3), 231–249 [vid. 2024-04-25]. ISSN 1469-7815, 0143-814X. Dostupné z: doi:10.1017/S0143814X00010035

AWAD, Mariette a Rahul KHANNA, 2015. Support Vector Machines for Classification. In: Mariette AWAD a Rahul KHANNA, ed. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers* [online]. Berkeley, CA: Apress, s. 39–66 [vid. 2024-07-10]. ISBN 978-1-4302-5990-9. Dostupné z: doi:10.1007/978-1-4302-5990-9_3

BARBERÁ, Pablo, Amber E. BOYDSTUN, Suzanna LINN, Ryan MCMAHON a Jonathan NAGLER, 2021. Automated Text Classification of News Articles: A Practical Guide. *Political Analysis* [online]. **29**(1), 19–42 [vid. 2024-04-29]. ISSN 1047-1987, 1476-4989. Dostupné z: doi:10.1017/pan.2020.8

BESTVATER, Samuel E. a Burt L. MONROE, 2023. Sentiment is Not Stance: Target-Aware Opinion Classification for Political Text Analysis. *Political Analysis* [online]. **31**(2), 235–256 [vid. 2024-04-26]. ISSN 1047-1987, 1476-4989. Dostupné z: doi:10.1017/pan.2022.10

BOOMGAARDEN, Hajo G., Claes H. DE VREESE, Andreas R.T. SCHUCK, Rachid AZROUT, Matthijs ELENBAAS, Joost H.P. VAN SPANJE a Rens VLIEGENTHART, 2013. Across time and space: Explaining variation in news coverage of the European Union. *European Journal of Political Research* [online]. **52**(5), 608–629 [vid. 2024-04-26]. ISSN 1475-6765. Dostupné z: doi:10.1111/1475-6765.12009

BOOMGAARDEN, Hajo G., Andreas R. T. SCHUCK, Matthijs ELENBAAS a Claes H. DE VREESE, 2011. Mapping EU attitudes: Conceptual and empirical dimensions of Euroscepticism and EU support. *European Union Politics* [online]. **12**(2), 241–266 [vid. 2024-04-25]. ISSN 1465-1165. Dostupné z: doi:10.1177/1465116510395411

BROSIUS, Anna, Erika J VAN ELSAS a Claes H DE VREESE, 2019. How media shape political trust: News coverage of immigration and its effects on trust in the European Union. *European Union Politics* [online]. **20**(3), 447–467 [vid. 2024-04-25]. ISSN 1465-1165. Dostupné z: doi:10.1177/1465116519841706

BROSIUS, Anna, Erika J VAN ELSAS a Claes H DE VREESE, 2020. Trust in context: National heuristics and survey context effects on political trust in the European Union. *European Union Politics* [online]. **21**(2), 294–311 [vid. 2024-04-26]. ISSN 1465-1165. Dostupné z: doi:10.1177/1465116519896010

CACCIATORE, Michael A., Dietram A. SCHEUFELE a Shanto IYENGAR, 2016. The End of Framing as we Know it ... and the Future of Media Effects. *Mass Communication and Society* [online]. **19**(1), 7–23 [vid. 2024-04-25]. ISSN 1520-5436. Dostupné z: doi:10.1080/15205436.2015.1068811

CAREY, Sean, 2002. Undivided Loyalties: Is National Identity an Obstacle to European Integration? *European Union Politics* [online]. **3**(4), 387–413 [vid. 2024-04-25]. ISSN 1465-1165. Dostupné z: doi:10.1177/1465116502003004001

CONNEAU, Alexis, Kartikay KHANDELWAL, Naman GOYAL, Vishrav CHAUDHARY, Guillaume WENZKE, Francisco GUZMÁN, Edouard GRAVE, Myle OTT, Luke ZETTLEMOYER a Veselin STOYANOV, 2020. *Unsupervised Cross-lingual Representation Learning at Scale* [online]. 7. duben 2020. B.m.: arXiv. [vid. 2024-07-11]. Dostupné z: doi:10.48550/arXiv.1911.02116

DE VRIES, Catherine E., 2018. *Euroscepticism and the Future of European Integration* [online]. B.m.: Oxford University Press [vid. 2024-04-25]. ISBN 978-0-19-183521-6. Dostupné z: doi:10.1093/oso/9780198793380.001.0001

DE WILDE, Pieter a Michael ZÜRN, 2012. Can the Politicization of European Integration be Reversed?*. *JCMS: Journal of Common Market Studies* [online]. **50**(s1), 137–153 [vid. 2024-04-25]. ISSN 1468-5965. Dostupné z: doi:10.1111/j.1468-5965.2011.02232.x

DE VREESE, Claes H. a Hajo G. BOOMGAARDEN, 2005. Projecting EU Referendums: Fear of Immigration and Support for European Integration. *European Union Politics* [online]. **6**(1), 59–82 [vid. 2024-04-25]. ISSN 1465-1165. Dostupné z: doi:10.1177/1465116505049608

ENTMAN, Robert M., 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication* [online]. **43**(4), 51–58 [vid. 2024-04-26]. ISSN 0021-9916. Dostupné z: doi:10.1111/j.1460-2466.1993.tb01304.x

GABEL, Matthew, 1998. Public Support for European Integration: An Empirical Test of Five Theories. *The Journal of Politics* [online]. **60**(2), 333–354 [vid. 2024-04-25]. ISSN 0022-3816. Dostupné z: doi:10.2307/2647912

GALPIN, Charlotte a Hans-Jörg TRENZ, 2019. Converging towards Euroscepticism? Negativity in news coverage during the 2014 European Parliament elections in Germany and the UK. *European Politics and Society* [online]. **20**(3), 260–276 [vid. 2024-04-26]. ISSN 2374-5118. Dostupné z: doi:10.1080/23745118.2018.1501899

GATTERMANN, Katjana a Claes DE VREESE, 2020. Awareness of Spitzenkandidaten in the 2019 European elections: The effects of news exposure in domestic campaign contexts. *Research & Politics* [online]. **7**(2), 2053168020915332 [vid. 2024-04-26]. ISSN 2053-1680. Dostupné z: doi:10.1177/2053168020915332

GRIMMER, Justin a Brandon M. STEWART, 2013. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* [online]. **21**(3), 267–297 [vid. 2024-04-30]. ISSN 1047-1987, 1476-4989. Dostupné z: doi:10.1093/pan/mps028

HANITZSCH, Thomas a Tim P VOS, 2018. Journalism beyond democracy: A new look into journalistic roles in political and everyday life. *Journalism* [online]. **19**(2), 146–164 [vid. 2024-04-25]. ISSN 1464-8849. Dostupné z: doi:10.1177/1464884916673386

HAVLÍK, Vlastimil, 2019. Technocratic Populism and Political Illiberalism in Central Europe. *Problems of Post-Communism* [online]. **66**(6), 369–384 [vid. 2024-04-26]. ISSN 1075-8216. Dostupné z: doi:10.1080/10758216.2019.1580590

HOBOLT, Sara B. a Catherine E. de VRIES, 2016. Public Support for European Integration. *Annual Review of Political Science* [online]. **19**(Volume 19, 2016), 413–432 [vid. 2024-04-25]. ISSN 1094-2939, 1545-1577. Dostupné z: doi:10.1146/annurev-polisci-042214-044157

HOOGHE, Liesbet, 2007. What Drives Euroskepticism?: Party-Public Cueing, Ideology and Strategic Opportunity. *European Union Politics* [online]. **8**(1), 5–12 [vid. 2024-04-24]. ISSN 1465-1165. Dostupné z: doi:10.1177/1465116507073283

HOOGHE, Liesbet a Gary MARKS, 2005. Calculation, Community and Cues: Public Opinion on European Integration. *European Union Politics* [online]. **6**(4), 419–443 [vid. 2024-04-25]. ISSN 1465-1165. Dostupné z: doi:10.1177/1465116505057816

HOOGHE, Liesbet a Gary MARKS, 2007. Sources of Euroscepticism. *Acta Politica* [online]. **42**(2), 119–127 [vid. 2024-04-25]. ISSN 1741-1416. Dostupné z: doi:10.1057/palgrave.ap.5500192

HOOGHE, Liesbet a Gary MARKS, 2009. A Postfunctionalist Theory of European Integration: From Permissive Consensus to Constraining Dissensus. *British Journal of Political Science* [online]. **39**(1), 1–23 [vid. 2024-04-25]. ISSN 1469-2112, 0007-1234. Dostupné z: doi:10.1017/S0007123408000409

HSIEH, Cho-Jui a Inderjit S DHILLON, 2014. A Divide-and-Conquer Solver for Kernel Support Vector Machines.

KHOO, Christopher SG a Sathik Basha JOHNKHAN, 2018. Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science* [online]. **44**(4), 491–511 [vid. 2024-04-26]. ISSN 0165-5515. Dostupné z: doi:10.1177/0165551517703514

KRIPPENDORFF, Klaus, 2004. *Content analysis: an introduction to its methodology*. 2nd ed. Thousand Oaks, Calif: Sage. ISBN 978-0-7619-1544-7.

KRZYŻANOWSKI, Michał, Anna TRIANDAFYLLIDOU a Ruth WODAK, 2018. The Mediatization and the Politicization of the “Refugee Crisis” in Europe. *Journal of Immigrant & Refugee Studies* [online]. **16**(1–2), 1–14 [vid. 2024-02-26]. ISSN 1556-2948. Dostupné z: doi:10.1080/15562948.2017.1353189

LECONTE, Cécile, 2015. From pathology to mainstream phenomenon: Reviewing the Euroscepticism debate in research and theory. *International Political Science Review* [online]. **36**(3), 250–263 [vid. 2024-04-26]. ISSN 0192-5121. Dostupné z: doi:10.1177/0192512115576236

LECHELER, Sophie a Claes H. DE VREESE, 2017. News Media, Knowledge, and Political Interest: Evidence of a Dual Role From a Field Experiment. *Journal of Communication* [online]. **67**(4), 545–564 [vid. 2024-04-26]. ISSN 0021-9916. Dostupné z: doi:10.1111/jcom.12314

LECHELER, Sophie a Claes H. de VREESE, 2018. *News Framing Effects* [online]. London: Routledge. ISBN 978-1-315-20807-7. Dostupné z: doi:10.4324/9781315208077

LIU, Bing, 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions* [online]. Cambridge: Cambridge University Press [vid. 2024-04-26]. Dostupné z: doi:10.1017/CBO9781139084789

MACEK, Jakub, Alena MACKOVÁ, Vassilis PAVLOPOULOS, Veronika KALMUS, C. Michael ELAVSKY a Jan ŠEREK, 2018. Trust in alternative and professional media: The case of the youth news audiences in three European countries. *European Journal of Developmental Psychology* [online]. **15**(3), 340–354 [vid. 2024-04-26]. ISSN 1740-5629. Dostupné z: doi:10.1080/17405629.2017.1398079

MARQUART, Franziska, Andreas C. GOLDBERG, Erika J. VAN ELSAS, Anna BROSIUS a Claes H. DE VREESE, 2019. Knowing is not loving: media effects on knowledge about and attitudes toward the EU. *Journal of European Integration* [online]. **41**(5), 641–655 [vid. 2024-04-25]. ISSN 0703-6337. Dostupné z: doi:10.1080/07036337.2018.1546302

MATTHES, Jörg a Matthias KOHRING, 2008. The Content Analysis of Media Frames: Toward Improving Reliability and Validity. *Journal of Communication* [online]. **58**(2), 258–279 [vid. 2024-04-30]. ISSN 1460-2466. Dostupné z: doi:e10.1111/j.1460-2466.2008.00384.x

MCCOMBS, Maxwell E., Donald L. SHAW a David H. WEAVER, 2014. New Directions in Agenda-Setting Theory and Research. *Mass Communication and Society* [online]. **17**(6), 781–802 [vid. 2024-04-25]. ISSN 1520-5436. Dostupné z: doi:10.1080/15205436.2014.964871

MOHAMMAD, Saif M., 2016. 9 - Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text. In: Herbert L. MEISELMAN, ed. *Emotion Measurement* [online]. B.m.: Woodhead Publishing, s. 201–237 [vid. 2024-04-26]. ISBN 978-0-08-100508-8. Dostupné z: doi:10.1016/B978-0-08-100508-8.00009-6

NFNZ.CZ, [b.r.]. MediaRating. *Nadační fond nezávislé žurnalistiky* [online] [vid. 2024-04-27]. Dostupné z: <https://www.nfnz.cz/rating-medii>

PAUL STATHAM a HANS-JÖRG TRENZ, 2012. *The Politicization of Europe: Contesting the Constitution in the Mass Media* [online]. B.m.: Routledge. 1st Edition [vid. 2024-04-25]. ISBN ISBN 9780415635660. Dostupné z: <https://www.routledge.com/The-Politicization-of-Europe-Contesting-the-Constitution-in-the-Mass-Media/Statham-Trenz/p/book/9780415635660>

PAWAR, Pratiksha Y. a S. H. GAWANDE, 2012. A Comparative Study on Different Types of Approaches to Text Categorization. *International Journal of Machine Learning and Computing* [online]. 423–426 [vid. 2024-07-25]. ISSN 20103700. Dostupné z: doi:10.7763/IJMLC.2012.V2.158

RAKUŠANOVÁ, Petra, 2007. The constitutional debate: A one man show? Vaclav Klaus and the constitutional discourse in the Czech Republic. *Perspectives on European Politics and Society* [online]. 8(3), 342–373 [vid. 2024-04-26]. ISSN 1570-5854. Dostupné z: doi:10.1080/15705850701433041

REIMERS, Nils a Iryna GUREVYCH, 2019. *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* [online]. 27. srpen 2019. B.m.: arXiv. [vid. 2024-04-30]. Dostupné z: doi:10.48550/arXiv.1908.10084

SALGADO, Susana, 2018. *Online media impact on politics. Views on post-truth politics and postpostmodernism* [online]. 2018. B.m.: Intellect. ISSN 2040-0918. Dostupné z: doi:https://doi.org/10.1386/macp.14.3.317_1

SÁNCHEZ-CUENCA, Ignacio, 2000. The Political Basis of Support for European Integration. *European Union Politics* [online]. 1(2), 147–171 [vid. 2024-04-25]. ISSN 1465-1165. Dostupné z: doi:10.1177/1465116500001002001

SEMETKO, Holli A. a Patti M. Valkenburg VALKENBURG, 2000. Framing European politics: A Content Analysis of Press and Television News. *Journal of Communication* [online]. 50(2), 93–109 [vid. 2024-04-26]. ISSN 0021-9916. Dostupné z: doi:10.1111/j.1460-2466.2000.tb02843.x

SHEHATA, Adam a Jesper STRÖMBÄCK, 2018. Learning Political News From Social Media: Network Media Logic and Current Affairs News Learning in a High-Choice Media Environment. *Communication Research* [online]. **48**(1), 125–147 [vid. 2024-04-25]. ISSN 0093-6502. Dostupné z: doi:10.1177/0093650217749354

STRÖMBÄCK, Jesper a Adam SHEHATA, 2019. The Reciprocal Effects Between Political Interest and TV News Revisited: Evidence From Four Panel Surveys. *Journalism & Mass Communication Quarterly* [online]. **96**(2), 473–496 [vid. 2024-04-25]. ISSN 1077-6990. Dostupné z: doi:10.1177/1077699018793998

TABOADA, Maite, 2016. Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics* [online]. **2**(Volume 2, 2016), 325–347 [vid. 2024-04-26]. ISSN 2333-9683, 2333-9691. Dostupné z: doi:10.1146/annurev-linguistics-011415-040518

TABOADA, Maite, Julian BROOKE, Milan TOFILOSKI, Kimberly VOLL a Manfred STEDE, 2011. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics* [online]. **37**(2), 267–307 [vid. 2024-04-26]. ISSN 0891-2017. Dostupné z: doi:10.1162/COLI_a_00049

VLIEGENTHART, Rens, Andreas R. T. SCHUCK, Hajo G. BOOMGAARDEN a Claes H. DE VREESE, 2008. News Coverage and Support for European Integration, 1990–2006. *International Journal of Public Opinion Research* [online]. **20**(4), 415–439 [vid. 2024-02-20]. ISSN 0954-2892. Dostupné z: doi:10.1093/ijpor/edn044

VREESE, Claes H., 2005. News framing: Theory and typology. *Information Design Journal* [online]. **13**(1), 51–62 [vid. 2024-04-26]. ISSN 0142-5471, 1569-979X. Dostupné z: doi:10.1075/idjdd.13.1.06vre

WALTER, Stefanie, 2017. Citizenship, Democracy and the EU. In: Stefanie WALTER, ed. *EU Citizens in the European Public Sphere: An Analysis of EU News in 27 EU Member States* [online]. Wiesbaden: Springer Fachmedien, s. 27–49 [vid. 2024-04-25]. ISBN 978-3-658-14486-9. Dostupné z: doi:10.1007/978-3-658-14486-9_2

ZHANG, Lei, Shuai WANG a Bing LIU, 2018. Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery* [online]. **8**(4), e1253 [vid. 2024-04-26]. ISSN 1942-4795. Dostupné z: doi:10.1002/widm.1253

ZULFIQAR, M., M. KAMRAN, M. B. RASHEED, T. ALQUTHAMI a A. H. MILYANI, 2022. Hyperparameter optimization of support vector machine using adaptive differential evolution for electricity load forecasting. *Energy Reports* [online]. **8**, 13333–13352 [vid. 2024-07-25]. ISSN 2352-4847. Dostupné z: doi:10.1016/j.egy.2022.09.188

SEZNAM TABULEK

Tabulka 1: Kódovací schéma.....	20
Tabulka 3: Rozložení sentimentu v anotovaných datech.....	23
Tabulka 4: Četnosti rámců v korpusu	25
Tabulka 5: Výsledky SVM modelu pro klasifikaci polaritu sentimentu s SBERT transformací	31
Tabulka 6: Výsledky SVM modelu pro klasifikaci rámců s SBETR transformací	31
Tabulka 7: Výsledky SVM modelu pro klasifikaci polaritu sentimentu s BoW transformací	33
Tabulka 8: Výsledky SVM modelu pro klasifikaci rámců s BoW transformací	34

SEZNAM GRAFŮ

Graf 1: Zastoupení médií v korpusu	18
Graf 2: Četnost rámců v korpusu.....	23
Graf 3: Revidované četnosti rámců v korpusu	24
Graf 4: Porovnání metrik pro klasifikaci polarity sentimentu a klasifikaci rámců	32
Graf 5: Přehled výkonnostních metrik pro SVM model s BoW transformací	34
Graf 6: Grafické srovnání metrik, při použití různých transformací textu	36

SEZNAM ZKRATEK

BERT – Bidirectional Encoder Representations from Transformers

BoW – bag of words

EU – Evropská Unie

NA – not applicable

např. – například

RBF – Radial Basis Function

SBERT – Sentence Bidirectional Encoder Representations from Transformers

SML – supervised machine learning

SVM – Support Vector Machines

tzv. – tak zvaně