

Bakalářská práce

# **Analýza textu s využitím aspektové analýzy sentimentu a pokročilých metod NLP**

**Vilém Fiala**

Fakulta informačních technologií  
Katedra aplikované matematiky  
Vedoucí: Ing. Lukáš Klein  
13. května 2025



## Zadání bakalářské práce

<b>Název:</b>	Analýza textu s využitím aspektové analýzy sentimentu a pokročilých metod NLP
<b>Student:</b>	Vilém Fiala
<b>Vedoucí:</b>	Ing. Lukáš Klein
<b>Studijní program:</b>	Informatika
<b>Obor / specializace:</b>	Umělá inteligence 2021
<b>Katedra:</b>	Katedra aplikované matematiky
<b>Platnost zadání:</b>	do konce letního semestru 2025/2026

### Pokyny pro vypracování

Cílem práce je prozkoumat a analyzovat textové dokumenty se zaměřením na cílenou (targeted) a aspektovou (aspect based) analýzu sentimentu, která umožní identifikaci a hodnocení názorů na konkrétní témata či entity v textu. Práce bude zahrnovat přehled současných přístupů využívajících machine learning a deep learning modely pro analýzu přirozeného jazyka (NLP), včetně porovnání vybraných metod a jejich výkonnosti na různých typech dat (jiný jazyk, jiný případ/téma). Jako zdroj dat bude využit rozsáhlý archiv Newton Media, který poskytne reálné příklady textů pro praktické experimenty.

Cíle:

1. Provedte přehled současných metod cílené a aspektové analýzy sentimentu.
2. Analyzujte a porovnejte výkonnost vybraných modelů strojového a hlubokého učení.
3. Implementujte a otestujte modely na reálných mediálních datech z archivu Newton Media.
4. Vyhodnoťte vhodnost metod pro český jazyk a mediální texty.

Doporučená metodologie:

Rešerše literatury o cílené a aspektové analýze sentimentu.

Předzpracování a anotace dat z archivu Newton Media.

Trénování modelů a jejich srovnání na různých typech textů.

České vysoké učení technické v Praze

Fakulta informačních technologií

© 2025 Vilém Fiala. Všechna práva vyhrazena.

*Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí a nad rámec oprávnění uvedených v Prohlášení, je nezbytný souhlas autora.*

Odkaz na tuto práci: Fiala Vilém. *Analýza textu s využitím aspektové analýzy sentimentu a pokročilých metod NLP*. Bakalářská práce. České vysoké učení technické v Praze, Fakulta informačních technologií, 2025.

*Velmi rád bych poděkoval vedoucímu mé bakalářské práce, Ing. Lukáši Kleinovi, za cenné rady, podnětnou zpětnou vazbu a kontinuální podporu během celého procesu psaní. Mé poděkování rovněž patří společnosti Newton Media za poskytnutí prostoru k vypracování bakalářské práce v rámci mého pracovního zařazení. Dále děkuji Ing. Petře Pavlíčkové, Ph.D., za cenné konzultace a ochotu zodpovídat všechny mé dotazy, jež mi zásadně pomohly při samotném psaní.*

## Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů. V souladu s ust. § 2373 odst. 2 zákona č. 89/2012 Sb., občanský zákoník, ve znění pozdějších předpisů, tímto uděluji nevýhradní oprávnění (licenci) k užití této mojí práce, a to včetně všech počítačových programů, jež jsou její součástí či přílohou a veškeré jejich dokumentace (dále souhrnně jen „Dílo“), a to všem osobám, které si přejí Dílo užít. Tyto osoby jsou oprávněny Dílo užít jakýmkoli způsobem, který nesnižuje hodnotu Díla a za jakýmkoli účelem (včetně užití k výdělečným účelům). Toto oprávnění je časově, teritoriálně i množstevně neomezené.

Prohlašuji, že jsem v průběhu příprav a psaní závěrečné práce použil nástroje umělé inteligence. Vygenerovaný obsah jsem ověřil. Stvrzuji, že jsem si vědom, že za obsah závěrečné práce plně zodpovídám.

V Praze dne 13. května 2025

## Abstrakt

Tato bakalářská práce se zaměřuje na problematiku aspektové analýzy sentimentu (ABSA) v různých jazycích a tematických oblastech. Nejprve představuje klíčové poznatky ze zpracování přirozeného jazyka (NLP) a analýzy sentimentu včetně vybraných přístupů a modelů. Následně porovnává jejich výkonnost v češtině i angličtině a zkoumá, jak se modely chovají při aplikaci na odlišné domény. Výsledkem je ucelený přehled o efektivitě různých modelů a doporučení nejvhodnějších postupů pro práci s českými texty, zejména v oblasti mediálních analýz.

**Klíčová slova** zpracování přirozeného jazyka, jazykové modely, předtrénované jazykové modely, aspektová analýza sentimentu, Python, BERT, Hugging Face

## Abstract

This bachelor's thesis focuses on the problem of aspect-based sentiment analysis (ABSA) across different languages and thematic domains. It first presents the key concepts of natural language processing (NLP) and sentiment analysis, including a discussion of selected approaches and models. Subsequently, it compares their performance in both Czech and English, examining how the models behave when applied to different domains. The thesis concludes with a comprehensive overview of the effectiveness of various models and recommendations for the most suitable methods when working with Czech texts, particularly in the context of media analysis.

**Keywords** natural language processing, language models, pretrained language models, aspect-based sentiment analysis, Python, BERT, Hugging Face

## Obsah

<b>Úvod</b>	<b>1</b>
Cíle	2
<b>1 Zpracování jazyka pomocí umělé inteligence</b>	<b>3</b>
1.1 Zpracování přirozeného jazyka – NLP	3
1.1.1 Dvě základní složky NLP	4
1.1.1.1 Pochopení přirozeného jazyka – NLU	4
1.1.2 Problematika vícejazyčného zpracování	5
1.1.3 Využití NLP v praxi	6
1.2 Jazykové modely – LM	7
1.2.1 Tokenizace	8
1.2.1.1 Tokeny v různých jazycích	8
1.2.2 Předtrénované jazykové modely – PLMs	8
1.2.2.1 Transformátorová architektura	9
1.2.2.2 Předtrénování	9
1.2.2.3 Fine-tuning	10
1.2.3 Decoder vs Encoder	10
1.2.3.1 Decoder-only modely	11
1.2.3.2 Encoder-only modely	11
1.2.3.3 Jaký model pro sentiment?	11
<b>2 Analýza sentimentu</b>	<b>12</b>
2.1 Techniky analýzy sentimentu	13
2.1.1 Slovníková metoda	13
2.1.2 Tradiční přístupy strojového učení	13
2.1.3 Hluboké učení	13
2.1.3.1 Konvoluční neuronové sítě	14
2.1.3.2 Rekurentní neuronové sítě	14
2.1.3.3 Sítě s dlouhou krátkodobou pamětí	14
2.1.3.4 Další metody hlubokého učení v NLP	15
2.1.4 Hybridní metoda	15
2.2 Úrovně analýzy sentimentu	15
2.2.1 Dokumentová úroveň	15
2.2.2 Větná úroveň	16
2.2.3 Aspektová úroveň	16
2.3 Aspektová analýza sentimentu	16

2.3.1	Základní dílčí úlohy ABSA . . . . .	17
2.3.2	Složené úlohy ABSA . . . . .	18
2.3.2.1	Extrakce dvojic . . . . .	18
2.3.2.2	Extrakce trojic . . . . .	19
2.3.2.3	Extrakce čtveřic . . . . .	19
2.3.3	Využití transformátorových modelů v ABSA . . . . .	20
2.3.3.1	Datasety . . . . .	20
2.3.3.2	Složené úlohy ABSA . . . . .	20
2.4	Využití analýzy sentimentu . . . . .	22
2.4.1	Obchodní oblast . . . . .	22
2.4.2	Doporučující systémy . . . . .	23
2.4.3	Zdravotnická oblast . . . . .	23
2.5	Otevřené problémy analýzy sentimentu . . . . .	24
<b>3</b>	<b>Český jazyk a analýza sentimentu</b>	<b>25</b>
3.1	Předešlé práce na analýzu sentimentu v češtině . . . . .	25
3.1.1	Problematika analýzy sentimentu v češtině . . . . .	25
3.1.2	Porovnání různých modelů . . . . .	26
3.1.3	Představení nových technik . . . . .	26
3.1.4	Datasety v češtině . . . . .	26
3.1.4.1	Analýza sentimentu na větné úrovni . . . . .	27
3.1.4.2	Aspektová analýza sentimentu . . . . .	27
3.1.5	České jazykové modely . . . . .	27
3.1.5.1	Český BERT – Czert . . . . .	27
3.1.5.2	Český RoBERTa – RobeCzech . . . . .	28
3.1.5.3	FERNET – BERT i RoBERTa . . . . .	28
3.2	Shrnutí dosavadního výzkumu . . . . .	28
<b>4</b>	<b>Metodologie a postup trénování</b>	<b>29</b>
4.1	Datasety . . . . .	29
4.1.1	Příprava datasetů . . . . .	30
4.1.2	Dataset SemEval-2014 . . . . .	30
4.1.2.1	Recenze notebooků . . . . .	31
4.1.2.2	Recenze restaurací . . . . .	32
4.1.3	Český dataset recenzí restaurací . . . . .	32
4.1.4	Data z mediální domény . . . . .	34
4.1.4.1	Popis generování dat . . . . .	35
4.1.4.2	Popis samotného datasetu . . . . .	35
4.2	Vybrané metody . . . . .	37
4.3	Vybrané modely . . . . .	38
4.3.1	Inovativní modely . . . . .	38
4.3.1.1	BERT . . . . .	38
4.3.1.2	mBERT . . . . .	39
4.3.1.3	RoBERTa . . . . .	40



4.3.1.4	XLM-RoBERTa . . . . .	40
4.3.1.5	DeBERTa-V3 . . . . .	40
4.3.1.6	ModernBERT . . . . .	41
4.3.2	Jazykově adaptované modely . . . . .	41
4.3.2.1	SlavicBERT . . . . .	42
4.3.2.2	Český BERT – Czert . . . . .	42
4.3.2.3	Český RoBERTa – RobeCzech . . . . .	42
4.3.2.4	Český BERT a RoBERTa – FERNET . . . . .	42
4.4	Popis přístupu . . . . .	43
4.4.1	Trénovací prostředí . . . . .	43
4.4.2	Použité knihovny . . . . .	43
4.4.3	Popis částí kódu . . . . .	44
4.4.3.1	Tokenizace . . . . .	44
4.4.3.2	Vyhodnocovací metrika . . . . .	45
4.4.3.3	Trénovací struktura . . . . .	45
4.4.3.4	Predikace sentimentu . . . . .	47
4.5	Vyhodnocovací metriky a praktické testy . . . . .	48
4.5.1	Vyhodnocovací metriky . . . . .	48
4.5.2	Praktické testy . . . . .	48
4.5.2.1	Recenze notebooků L . . . . .	49
4.5.2.2	Recenze restaurací R . . . . .	49
4.5.2.3	Mediální texty M . . . . .	50
4.5.2.4	Obecná doména O . . . . .	50
<b>5</b>	<b>Výsledky a porovnání</b>	<b>52</b>
5.1	Výsledky . . . . .	52
5.1.1	Recenze notebooků – angličtina . . . . .	53
5.1.2	Recenze restaurací – angličtina . . . . .	54
5.1.3	Recenze restaurací – čeština . . . . .	55
5.1.4	Mediální články – čeština . . . . .	56
5.2	Porovnání použitých modelů a metod . . . . .	57
5.2.1	Porovnání metod – Klasická vs. QA-M . . . . .	57
5.2.2	Porovnání modelů na anglických datech . . . . .	58
5.2.3	Porovnání modelů na různých jazycích . . . . .	59
5.2.4	Porovnání modelů na českých datech . . . . .	59
5.3	Shrnutí a doporučení . . . . .	60
5.3.1	Anglické datasety . . . . .	60
5.3.2	České datasety . . . . .	61
5.3.3	Doporučení pro české mediální texty . . . . .	61
5.3.4	Analýza sentimentu pomocí LLM . . . . .	61
5.3.4.1	Popis modelů a použitého promptu . . . . .	62
5.3.4.2	Výsledky a závěr k LLM . . . . .	62
5.4	Další výzkum . . . . .	63

Obsah	ix
<b>Závěr</b>	<b>64</b>
<b>A Tabulky výsledků trénování a praktických testů</b>	<b>65</b>
<b>Obsah příloh</b>	<b>99</b>

## Seznam obrázků

2.1	Jednotlivé dílčí úlohy aspektové analýzy sentimentu . . . . .	17
4.1	Rozdělení sentimentu v datasetech . . . . .	30
4.2	Rozdělení délky textů v datasetu Recenze notebooků v angličtině	31
4.3	Porovnání dvou tokenizátorů na datasetu Recenze notebooků v angličtině . . . . .	32
4.4	Rozdělení délky textů v datasetu Recenze restaurací v angličtině	33
4.5	Porovnání dvou tokenizátorů na datasetu Recenze restaurací v angličtině . . . . .	33
4.6	Rozdělení délky textů v datasetu Recenze restaurací v češtině .	34
4.7	Porovnání dvou tokenizátorů na datasetu Recenze restaurací v češtině . . . . .	34
4.8	Rozdělení délky textů v datasetu Mediální texty v češtině . . .	36
4.9	Porovnání dvou tokenizátorů na datasetu Mediální texty v češtině	36

## Seznam tabulek

4.1	Parametry použitých modelů . . . . .	39
5.1	Recenze notebooků v angličtině – malá tabulka . . . . .	53
5.2	Recenze restaurací v angličtině – malá tabulka . . . . .	54
5.3	Recenze restaurací v češtině – malá tabulka . . . . .	55
5.4	Mediální články v češtině – malá tabulka . . . . .	56
A.1	Recenze notebooků v angličtině – trénování 1 . . . . .	66
A.2	Recenze notebooků v angličtině – trénování 2 . . . . .	67
A.3	Recenze notebooků v angličtině – praktické testy 1 . . . . .	68
A.4	Recenze notebooků v angličtině – praktické testy 2 . . . . .	69
A.5	Recenze restaurací v angličtině – trénování 1 . . . . .	70
A.6	Recenze restaurací v angličtině – trénování 2 . . . . .	71
A.7	Recenze restaurací v angličtině – praktické testy 1 . . . . .	72
A.8	Recenze restaurací v angličtině – praktické testy 2 . . . . .	73
A.9	Recenze restaurací v češtině – trénování 1 . . . . .	74
A.10	Recenze restaurací v češtině – trénování 2 . . . . .	75
A.11	Recenze restaurací v češtině – praktické testy 1 . . . . .	76
A.12	Recenze restaurací v češtině – praktické testy 2 . . . . .	77
A.13	Mediální texty v češtině – trénování 1 . . . . .	78
A.14	Mediální texty v češtině – trénování 2 . . . . .	79
A.15	Mediální texty v češtině – praktické testy 1 . . . . .	80
A.16	Mediální texty v češtině – praktické testy 2 . . . . .	81

## Seznam výpisů kódu

4.1	Ukázka promptu pro extrakci nejdůležitějších aspektů . . . . .	35
4.2	Ukázka promptu pro generaci shrnutí . . . . .	35
4.3	Ukázka funkce tokenizace u základní metody . . . . .	45
4.4	Ukázka funkce tokenizace u metody QA-M . . . . .	45
4.5	Ukázka funkce pro výpočet přesnosti . . . . .	46
4.6	Ukázka struktury TrainingArguments . . . . .	46
4.7	Ukázka struktury Trainer . . . . .	47
4.8	Ukázka funkce na predikci sentimentu . . . . .	47
5.1	Ukázka promptu v češtině pro analýzu sentimentu . . . . .	62
5.2	Ukázka promptu v angličtině pro analýzu sentimentu . . . . .	62

## Seznam zkratek

ABSA	Aspect-based sentiment analysis
ACD	Aspect category detection
ACOS	Aspect-Category-Opinion-Sentiment Quadruple Extraction
ACSA	Aspect Category Sentiment Analysis
ALBERT	A Lite BERT
AOPE	Aspect-Opinion Pair Extraction
ASC	Aspect sentiment classification
ASTE	Aspect Sentiment Triplet Extraction
ATE	Aspect term extraction
BERT	Bidirectional Encoder Representations from Transformers
BPE	Byte-pair encoding
CLM	Conventional language models
CNN	Convolutional neural networks
DeBERTa	Decoding-enhanced BERT with Disentangled Attention
E2E-ABSA	End-to-End ABSA
FERNET	Flexible Embedding Representation NETwork
GPT	Generative Pre-training Transformer
LLM	Large language model
LM	Language model
LSTM	Long Short Term Memory
MAMS	Multi-Aspect Multi-Sentiment
mBERT	Multilingual BERT
MemN2N	End-to-End Memory Network
MemNN	Memory networks
NER	Named entity recognition
NB	Naive Bayes
NLG	Natural language generation
NLP	Natural language processing
NLU	Natural language understanding
OATS	Opinion Aspect Target Sentiment
OPE	Opinion term extraction
PLM	Pre-trained language models
RNN	Recurrent neural networks
RoBERTa	Robustly Optimized BERT Pretraining Approach
SRL	Semantic Role Labeling
SVM	Support vector machines

TASD	Target Aspect Sentiment Detection
UABSA	Unified ABSA
XLM	Cross-lingual language models

# Úvod

Analýza médií a textů je díky rychlému rozvoji umělé inteligence podstatně jednodušší než dříve. Firmy již nemusí zaměstnávat početné týmy pracovníků, kteří by osobně četli každý článek a na základě subjektivních pocitů jej hodnotili. Takové řešení navíc často přináší nekonzistentní výsledky, jelikož každý člověk může k hodnocení přistupovat odlišně. Proto se v současnosti stále více organizací obrací k umělé inteligenci, která – po svém natrénování – poskytuje konzistentní hodnocení.

Ne vždy je přitom nutné provádět detailní rozbor každého textu, mnohdy postačí obecná informace o celkovém sentimentu článku či recenze. Pro firmy nabízející široké portfolio produktů bývá důležité zejména to, jaký podíl recenzí je pozitivních a jaký negativních, což jim poskytuje rychlý přehled o vnímání jejich zboží ze strany zákazníků.

Problém však nastává, pokud jedna recenze či článek obsahuje kladné i záporné hodnocení. Například u notebooku si může autor recenze pochvalovat kvalitní obrazovku a dlouhou výdrž baterie, ale současně kritizovat nedostatečnou odolnost klávesnice. V takové situaci je vhodnější zaměřit se na konkrétní aspekty či témata a zjišťovat, jaký postoj k nim text zaujímá.

Cílený sentiment proto představuje užitečnou alternativu, kdy se text analyzuje s ohledem na zvolený aspekt a výsledkem je hodnocení vztahující se k tomuto aspektu. Tento postup umožňuje rychle identifikovat názor na vybrané aspekty a efektivně tak třídit či porovnávat texty podle konkrétních kritérií.

Zároveň však dnes neustále vznikají nové modely a metody, což klade nároky na uživatele, kteří musejí testovat různé přístupy a hledat nejvhodnější řešení pro svůj konkrétní problém. Analýza sentimentu v českém jazyce je navíc v porovnání s angličtinou stále poměrně málo rozšířená, což komplikuje výběr vhodného modelu a metod pro české texty a recenze.



## Cíle

Cílem této bakalářské práce je prozkoumat současné metody aspektové analýzy sentimentu v textu a otestovat vybrané postupy na různých typech dat v různých jazycích. Teoretická část bude zaměřena na vymezení problematiky zpracování přirozeného jazyka, analýzy sentimentu a popis vybraných metod i postupů. V praktické části budou vybrané metody aplikovány na data v češtině a v angličtině, vzájemně porovnány a vyhodnoceny z hlediska vhodnosti pro daný jazyk a konkrétní téma. Smyslem je najít či doporučit nejvhodnější metody pro práci s českými texty v kontextu mediálních textů.

# Zpracování jazyka pomocí umělé inteligence

Umělá inteligence pronikla do téměř všech oblastí lidské činnosti – od generování a úprav multimediálního obsahu až po sofistikované predikční úlohy. Jednou z nejviditelnějších aplikací je v posledních letech zpracování textu, tedy schopnost strojově analyzovat a vytvářet přirozený jazyk. Významný rozvoj výpočetních kapacit a metod strojového učení umožnil vznik rozsáhlých modelů, které v řadě lingvistických úloh dosahují až lidské úrovně.

Tato kapitola nejprve představí základní principy zpracování přirozeného jazyka (NLP) a vymezí jeho roli v rámci širšího pole umělé inteligence. Následně se zaměří na jazykové modely jako klíčový nástroj praktického využití NLP. Popisuje jejich architektury i mechanismy učení, přičemž zvláštní pozornost je věnována modelům využitým v dalších částech této práce.

## 1.1 Zpracování přirozeného jazyka – NLP

*Natural language processing* (NLP), v češtině zpracování přirozeného jazyka, spojuje poznatky z lingvistiky a umělé inteligence s cílem umožnit počítačům porozumět slovům a výrazům, jež lidé používají v běžném jazyce. [1] Programovací jazyky (např. C, C++ či Python) byly navrženy, aby programátoři mohli počítačům předávat přesné pokyny. Pro takovou komunikaci však musí člověk nejprve detailně zvládnout syntaxi a pravidla jazyka – a i pak může drobná syntaktická či pravopisná chyba zabránit správnému provedení příkazu.

S přirozeným jazykem je to ale složitější. Když se v běžné komunikaci udělá menší chyba, často si druhá strana s pomocí kontextu domyslí, co bylo myšleno. Pro počítač je takové „domýšlení“ mnohem obtížnější. Právě to je jeden z úkolů, jimiž se NLP zabývá: vyvinout modely či algoritmy, které dokážou porozumět lidské řeči (byť s chybami) a provést zadané pokyny.

### 1.1.1 Dvě základní složky NLP

NLP se obvykle dělí na dvě hlavní části:

1. *Natural Language Understanding* (NLU) – pochopení přirozeného jazyka.
2. *Natural Language Generation* (NLG) – generování přirozeného jazyka.

NLU umožňuje počítačům porozumět významu vět a slov, zachytit klíčové informace (např. emoce, zmíněné entity, hlavní témata) a rozpoznat, co je podstatou sdělení. NLG se naopak soustředí na tvorbu gramaticky a významově správných vět či textů. [1] Tato práce se zaměřuje především na NLU, protože při analýze sentimentu – tedy při zjišťování, jaký postoj daný text vyjadřuje – se primárně texty zpracovávají a vyhodnocují, aniž by musely být generovány.

#### 1.1.1.1 Pochopení přirozeného jazyka – NLU

Jak již bylo zmíněno, cílem NLU je naučit počítač porozumět psanému textu v přirozeném jazyce. K tomu se využívají poznatky z lingvistiky, která se dále dělí na několik rovin; každá zachycuje odlišný aspekt jazyka a společně popisují informace v něm obsažené. Níže je uveden jejich stručný přehled převzatý z díla [1]:

**Fonologická rovina** Zabývá se zvukovou stránkou jazyka. Zkoumá, jak jsou zvuky uspořádány a jak ovlivňují význam v komunikaci. V praxi se tak sleduje, jak se hlásky v jazyce kombinují a jak rozdílné zvukové podoby mohou měnit či vyjadřovat význam.

**Morfologická rovina** Zkoumá vnitřní strukturu slov a jejich nejmenší významové jednotky, tzv. morfémy. Každé slovo může být rozloženo na různé morfémy (např. předpony, kořen, přípony), které společně nesou určitý význam. Morfémy mají také různé funkce, např. předpony mohou být slovotvorné nebo tvarotvorné. [2] Tímto způsobem lze pochopit a analyzovat význam, který se skrývá už v samotné struktuře slova.

**Lexikální rovina** Zaměřuje se na význam jednotlivých slov a jejich přiřazení k odpovídajícím slovním druhům (*part-of-speech* tag). Pokud se jedno slovo může v různých kontextech chovat například jako podstatné jméno i sloveso, je na základě okolního textu určeno, který slovní druh v dané situaci plní. Součástí analýzy je také dělení textu na věty a slova, odstranění stop slov (např. „a“, „ale“, „že“), a využití technik *stemming* a *lemmatizace*.

- **Stemming** odstraňuje koncové části slov a redukuje je na jejich základní kořen. Například slova „vařený“ nebo „vařit“ by se mohla stáhnout ke tvaru „vař“.

- **Lemmatizace** pracuje s jazykovými slovníky, aby identifikovala správný tvar slova včetně ohýbání či změn rodů a pádů (u slova „vařil“ či „vařením“ se hledá výchozí tvar „vařit“).

**Syntaktická rovina** Zkoumá gramatickou strukturu vět a určuje vztahy mezi slovy. Po přiřazení slovních druhů na lexikální úrovni se slova seskupují do frází, které pak tvoří věty (*parsing*). Tento postup sleduje správné pořadí slov a vazby mezi nimi. Na rozdíl od předchozí úrovně neodstraňuje stop slova ani neužívá lemmatizaci či stemming, protože by tím mohl pozměnit gramatický význam.

**Sémantická rovina** Soustředí se na odhalení skutečného významu věty a na rozlišení možných interpretací. Zohledňuje nejen slovní zásobu, ale také kontext a vztahy mezi slovy. Díky tomu lze poznat, že text pojednává například o filmu, i když výslovně neobsahuje slovo „film“ (stačí pojmy jako „herec“, „režisér“ nebo „scénář“). Součástí této úrovně je i řešení slov s více významy.

**Diskurzí rovina** Zabývá se významovými souvislostmi napříč vícero větami a zajišťuje tak celkovou návaznost textu. Zaměřuje se především na vztahy mezi výrazy, aby bylo zřejmé, že určitý zájmeno či jiný jazykový prvek odkazuje na předchozí jméno nebo entitu. Tato analýza je důležitá pro pokročilé úlohy, jako je automatická sumarizace či extrakce informací, protože pomáhá správně určit, kdo nebo co je v textu zmíněno.

**Pragmatická rovina** Zkoumá, jaký vliv na porozumění má kontext a reálné znalosti o světě, jež nejsou přímo obsaženy v textu. Soustředí se na to, co mluvčí skutečně zamýšlí říct a jak posluchač dané sdělení interpretuje na základě širších souvislostí. Díky pragmatické analýze lze rozlišit například, zda otázka „Víte, kolik je hodin?“ slouží k zjištění času, nebo spíše vyjadřuje nespokojenost se zpožděním. Tato rovina tak doplňuje čistě jazykové rozbory o význam, který se odvozuje i z mimojazykových faktorů. [1]

Cílem zpracování přirozeného jazyka (NLP) je často sladit několik různých rovin tak, aby algoritmus či systém dokázal efektivně porozumět textu. Velkou výzvou přitom často bývá nejen kombinace více rovin, ale i řešení nejednoznačností, které se nejčastěji objevují v syntaktické, sémantické či lexikální rovině. [1]

### 1.1.2 Problematika vícejazyčného zpracování

Při zpracování přirozeného jazyka vychází každá lingvistická rovina z odlišných pravidel a charakteristik konkrétního jazyka. Rozdíly se týkají například

abeced, směru psaní, výslovnosti, gramatiky, tvorby slov či struktury vět. Neexistuje proto jediné univerzální řešení, které by fungovalo stejně dobře pro všechny jazyky.

Tato práce se zaměřuje zejména na češtinu a angličtinu. Vzhledem k obrovskému rozšíření angličtiny – podle statistik W3Techs přes 49 % webových stránek používá angličtinu jako základní jazyk (oproti 1 % v češtině) [3] – je pro ni dostupné výrazně větší množství dat a zdrojů k trénování modelů. V českém prostředí je naopak často obtížnější získat dostatečně rozsáhlé a kvalitní datasey, což může omezovat přesnost metod vyvíjených pro češtinu.

### 1.1.3 Využití NLP v praxi

V předchozích částech byly popsány vlastnosti NLP a principy jeho jazykové analýzy. Metody NLP nacházejí široké uplatnění a dokážou zefektivnit celou řadu úkolů. Níže jsou uvedeny některé příklady, které tato práce dále podrobněji nerozebírá. Analýza sentimentu, na niž je tato práce soustředěna, bude předmětem kapitoly 2. Jednotlivé typy využití byly též zmíněny v díle [1]:

**Strojový překlad** Umožňuje automatické převádění textů z jednoho jazyka do druhého. Kromě samotného překladu slov se klade důraz na zachování významu, gramatických vazeb i kontextu, aby byl výsledný text srozumitelný a co nejbližší originálu.

**Kategorizace textu** Slouží k automatickému třídění rozsáhlých množství dokumentů (např. zprávy, oficiální spisy, tržní data) do předem definovaných kategorií. Urychluje vyhledávání a usnadňuje následné zpracování velkých datových souborů.

**Extrakce informací** Cílem je rozpoznat a vytáhnout důležité prvky z textu, například jména osob, místa či události. Tato informace se dále využívá pro sumarizaci, budování databází nebo jako vstup pro další analýzy.

**Sumarizace** Z textu se automaticky vytváří kratší přehled, který vystihuje klíčové informace. Umožňuje rychlé seznámení s obsahem rozsáhlých dokumentů či vícero zdrojů najednou.

**Dialogové systémy** Zahrnují chatovací asistenty nebo hlasové pomocníky. Na základě jazykové analýzy rozumějí dotazům uživatele a generují odpovědi nebo vykonávají zadané úkoly (například vyhledání informací či nastavení budíku). [1]

## 1.2 Jazykové modely – LM

V předchozí sekci 1.1 bylo naznačeno, jakým způsobem může být poskytnuta možnost pro běžného uživatele komunikovat se systémem v přirozeném jazyce, a jak se počítač pokouší porozumět zadaným pokynům. Nastává však otázka, jakým způsobem takovou komunikaci reálně implementovat a jak naprogramovat modely, které dokážou zpracovávat a analyzovat text prostřednictvím NLP. Tato problematika je řešena pomocí LM (*Language models*, česky jazykové modely).

Pojem *jazykové modely* je velmi obecný a zahrnuje řadu přístupů s různou hloubkou i rozsahem. Starší definice často zmiňují nástroj, který na základě rozsáhlé databáze textů přiřazuje větám či slovům určitou pravděpodobnost. Jde tedy o systém, jenž vychází z trénovacího korpusu a následně posuzuje, jak „pravděpodobná“ je daná sekvence slov. Lze si jej představit jako komplexní vyhledávací databázi, která na základě podobností generuje možné textové výstupy. [4]

Moderní přístupy staví zejména na neuronových sítích. Ty se po zpracování velkého množství dat naučí strukturu a generování textu efektivněji než tradiční statistické metody. [5] Přesto zůstává základní myšlenka stejná: cílem je naučit se pravděpodobnostní rozdělení jazykových jednotek (například slov a vět), a to tak, aby model dokázal předvídat či tvořit text na základě naučeného kontextu.

Tyto dvě definice společně ukazují dva hlavní přístupy, jak k tvorbě jazykových modelů lze přistupovat: [5, 6]

- **Statistický (klasický) přístup** – založený na myšlence, že pravděpodobnost určitého slova závisí na jeho předchozím kontextu. Cílem je sestavit pravděpodobnostní rozdělení nad posloupnostmi slov pomocí četností v trénovacím korpusu. Mezi známé statistické přístupy patří např. N-gram modely [7].
- **Neuronový (data-driven) přístup** – vychází z trénování neuronové sítě. Model se učí přímo z velkých objemů neupravených dat, čímž si buduje bohatou představu o slovech, frázích i kontextu. Mezi známe patří GPT [8] a BERT [9].

Toto rozdělení nás přivádí k rozlišení CLMs (*Conventional Language Models*, česky konvenční či klasické jazykové modely) a PLMs (*Pre-trained Language Models*, česky předtrénované jazykové modely) [5]. Tato práce se zaměří na PLMs, jelikož klasické statistické metody (CLMs) se dnes téměř nepoužívají a zároveň je velmi snadné PLMs implementovat či dotrénovat pomocí dostupných nástrojů, například prostřednictvím webového portálu Hugging Face [10].

Než budou představeny PLMs, je nejprve nutné text vhodně předzpracovat. Počítače totiž dokážou efektivněji ukládat slova, než aby si pamatovaly každý jednotlivý znak. K tomuto účelu slouží tokenizace.

### 1.2.1 Tokenizace

Než model začne zpracovávat text a učit se jeho strukturu, je třeba rozhodnout, v jaké formě bude tento text zpracovávat. Měl by text chápat jako lidé, tedy rozdělovat ho na slova, nebo i na víceslovné výrazy? A nebo jej rozdělit kompletně na jednotlivá písmena? To jsou otázky, kterými se vědci a programátoři zabývají již dlouho (viz studie z roku 1992 [11]).

Donedávna většina modelů pro NLP používala slova jako hlavní jednotku textu, ale s příchodem byte-pair encoding (BPE) se přešlo na části slov, tedy na *tokeny*. Díky tokenům se daly zmenšit slovníky modelů, aniž by došlo ke ztrátě důležitých informací. Tokeny mají více definic, ale obecně lze říci, že se jedná o tzv. *typografickou jednotku*, která je pro model nejlogičtější pro zpracování [12].

#### 1.2.1.1 Tokeny v různých jazycích

Jak již bylo naznačeno v podsekcí 1.1.2, každý jazyk má odlišnou gramatiku a pravidla, což se odráží i na způsobu tokenizace. Některé jazyky je možné snadněji převést na tokeny, zatímco u jiných bude potřeba více tokenů pro zachování významu jednoho slova.

Například záleží, zda jazyk obsahuje pády, či nikoli. V jazycích s pády, jako je čeština, jedno slovo může být reprezentováno více tokeny v závislosti na kontextu. Jednoduchým příkladem je slovo „ty“ v češtině a „you“ v angličtině. V angličtině zůstává „you“ vždy stejné, zatímco v češtině bude záviset na pádu – může to být například ve 2. pádu „tebe“, ve 3. pádu „tobě“ a tak dále. Tento faktor způsobuje, že slovník tokenů pro češtinu bude mnohem větší než pro angličtinu.

V článku [13] autorka zkoumá, jak funguje tokenizace v různých jazycích. Jako příklad je uvedena věta: „Jaké bude počasí příští týden“, která byla analyzována v pěti různých jazycích. V angličtině a španělštině měla tato věta 7, respektive 8 tokenů, zatímco v barmštině (jazyk v Myanmaru) a amharštině (jazyk v Etiopii) měla 61, respektive 69 tokenů, což je téměř desetkrát více. Podrobnější popis této problematiky je uveden v již zmíněném článku [13].

### 1.2.2 Předtrénované jazykové modely – PLMs

Předtrénované jazykové modely patří v současné době mezi nejvyužívanější nástroje v oblasti NLP. Jedná se o rozsáhlé neuronové sítě založené na transformátorové architektuře, které lze nasadit na širokou škálu úloh. Velká část jejich schopností vychází z faktu, že jsou nejprve natrénovány na rozsáhlých a různorodých korpusech, kde si osvojují znalost syntaxe, sémantiky a dalších jazykových jevů (viz též 1.1.1.1). Poté lze modely jemně doladit (tzv. *fine-tuning*) pro specifické použití. [5, 6, 14, 15]

V této práci bude takový typ modelů využit pro úlohu rozpoznávání sentimentu. Díky již nabyté znalosti struktury textu v češtině i angličtině postačí

model následně specializovaně dotrénovat tak, aby přesně klasifikoval náladu či postoj vyjádřený v daném textu, a to na základě vybraného trénovacího datasetu.

V následujících podsekcích jsou popsány stěžejní vlastnosti předtrénovaných jazykových modelů (PLM). Nejprve je představena jejich architektura, poté je přiblížen postup samotného předtrénování a nakonec je ukázáno, jak lze tyto modely doladit (*fine-tuning*) pro konkrétní úlohy.

### 1.2.2.1 Transformátorová architektura

Transformátor [16] představuje vylepšenou architekturu oproti rekurentním a konvolučním neuronovým sítím a postupně je nahradil v mnoha úlohách. Místo rekurence pracuje pouze s *mechanismem pozornosti* (attention mechanism), který dokáže zachytit globální závislosti mezi vstupem a výstupem. Díky tomu je velmi flexibilní (tzv. vysoce parametrizovatelná) a umožňuje vznik pokročilých modelů, jako jsou například GPT [8] nebo BERT [9].

Funkci pozornosti lze popsat jako mapování dotazu a sady klíč–hodnota na výstup, který je váženým součtem hodnot, váhy se odvozují z porovnání dotazu s příslušným klíčem. Namísto jedné takovéto pozornosti používá transformátor tzv. *multi-head attention*, jež poskytuje bohatší pohled na reprezentaci dat. Detailní popis *multi-head attention* i celé transformátorové architektury lze nalézt v původní práci [16].

### 1.2.2.2 Předtrénování

Předtrénování je klíčovou součástí moderní přípravy modelů. Dnes už není nutné trénovat každý model od začátku; místo toho stačí využít předtrénovaný model, který si lze stáhnout z dostupných portálů (například již zmíněného Hugging Face [10]), a následně ho dotrénovat na specifický úkol pomocí menšího datasetu. Díky tomu, že byly modely natrénovány na rozsáhlých a různorodých datech, dokážou zachytit nuance v různých jazycích a efektivně reagovat na úkoly.

Existuje mnoho metod předtrénování, které závisí na konkrétní úloze, pro kterou je model připravován. Nejznámější a nejčastěji používanou metodou je *missing token prediction* (predikce chybějícího tokenu). Kromě této metody existuje i řada dalších přístupů, jako např. *next-sentence prediction* (predikce následující věty) nebo *replaced token prediction* (predikce vyměněného tokenu). [5, 6]

**Missing token prediction** Tento typ trénování využívají jak autoregresní jazykové modely (např. GPT [8]), tak i maskované jazykové modely (např. BERT [9]). Cílem je pomocí určitých tokenů (slov) predikovat okolní nebo následující tokeny. Maskované jazykové modely jsou trénovány tak, že se přibližně 15% tokenů zamaskuje a pomocí zbytku se model snaží tyto maskované tokeny



predikovat. U autoregresivních jazykových modelů se na základě dané sekvence tokenů predikuje následující token v řadě. Hlavní rozdíl spočívá v tom, že první typ modelu predikuje tokeny v závislosti na všech dostupných, což zlepšuje celkový přehled o struktuře textu. Druhý přístup je vhodnější pro generaci slov, protože na základě určitých tokenů dokáže model vygenerovat zbytek textu. [5, 14]

**Další metody předtrénování** Jak již bylo zmíněno, existuje řada dalších metod předtrénování. *Next-sentence prediction* vybírá dvě věty a predikuje, zda druhá věta skutečně následuje po té první, čímž ověřuje správnost jejich pořadí. *Replaced token prediction* se zaměřuje na kontrolu, zda byly některé tokeny v textu vyměněny. Podrobnější popis těchto metod a dalších je k nalezení v těchto studiích [5, 6].

### 1.2.2.3 Fine-tuning

V předchozí podsekcí 1.2.2.2 bylo popsáno, proč je předtrénování klíčové a jak model získává základní znalost jazyka. Fine-tuning (dotrénování) tento již existující model přizpůsobuje konkrétnímu úkolu, například klasifikaci sentimentu nebo rozpoznávání pojmenovaných entit.

Nejběžnější přístup spočívá v tom, že se celý model znovu trénuje (resp. aktualizují se všechny jeho vrstvy) na menším datasetu určeném pro danou úlohu. Nevýhodou takového postupu ale je, že i při malé změně úkolu je nutné trénovat model od začátku, což je výpočetně náročné. [5, 14]

Proto se často používá tzv. *Adapter tuning* [17, 18], při němž se do modelu přidají speciální vrstvy (adapters) s relativně malým počtem parametrů. Během trénování se upravují pouze tyto adaptéry, zatímco původní parametry modelu zůstávají beze změny. Díky tomu lze jeden předtrénovaný model efektivně použít pro více odlišných úkolů, neboť stačí vyměnit jen konkrétní adaptéry místo kompletního přeučování celého modelu. [5, 14] Pro rozsáhlé modely jako GPT-3, který má 175 miliard parametrů [19], je tento přístup časově i výpočetně výrazně úspornější.

### 1.2.3 Decoder vs Encoder

V podsekcí 1.2.2.2 byly představeny dva rozdílné přístupy k trénování jazykových modelů. Modely trénované prvním přístupem jsou označovány jako *autoregresivní jazykové modely*, zatímco druhým jako *maskované jazykové modely*. Toto rozdělení vede k modelům *Decoder-only* a *Encoder-only*, které se liší nejen v přístupu k trénování, ale i v architektuře transformátoru. Architektura transformátoru obsahuje strukturu *encoder-decoder*, kterou lze rozdělit na dvě samostatné komponenty [16]. Tato možnost vedla k vytvoření modelů jako je GPT [8], které využívají pouze decoder, a BERT [9], který využívá pouze encoder.

### 1.2.3.1 Decoder-only modely

Díky velkému nárůstu zájmu o LLMs [20, 21] (*Large language models*, česky velké jazykové modely) se v současnosti věnuje většina pozornosti decoder-only neboli *generativním modelům*. Ty dokážou generovat texty působící jako lidské a zvládají interaktivní konverzaci. Na encoder-only modely se přitom často zapomíná. [22]

Současné generativní modely jsou však obvykle velmi rozsáhlé, pomalé a nákladné pro většinu úkolů. Přestože zvládnou všechno, co encoder-only modely a ještě více, existují situace, kdy je potřeba zpracovávat úlohy rychle a ve velkém množství. V takových případech je vhodnější menší model s nižšími nároky na výpočetní zdroje. Proto mají encoder-only modely stále své místo a praktické využití. [22]

### 1.2.3.2 Encoder-only modely

Výstup encoder-only modelů není text (jako u decoder-only modelů), nýbrž seznam hodnot (tzv. *embedding vector*). Místo generování slov tedy model svou odpověď „zakóduje“ do vektorové reprezentace, jež lze následně interpretovat nebo použít v jiné nadstavbové úloze. Proto se encoder-only modely někdy označují jako *reprezentační modely* [22].

Jak již bylo zmíněno, decoder-only modely mohou v principu plnit stejnou funkci jako encoder-only modely, avšak při čistě analytických úkolech může být jejich autoregresivní charakter nevýhodou. Encoder-only modely, díky jejich architektuře, jsou vhodnější pro rychlé a rozsáhlé analýzy textu, ať už jde o klasifikaci, detekci entit nebo jiné úlohy. Další přednosti encoder-only modelů jsou podrobně rozebrány například v blogu o modelu ModernBERT [22].

### 1.2.3.3 Jaký model pro sentiment?

Jelikož je v této práci hlavním cílem analyzovat sentiment textu, není nutná jeho generace. Stačí tedy číselná reprezentace, která vyjádří, jaký sentiment daný text má. Z tohoto důvodu jsou v této práci využity *encoder-only* modely, konkrétně BERT [9] a jeho následné varianty, jako např. RoBERTa [23] či ModernBERT [24].

[illegible][illegible][illegible][illegible][illegible]

video a text), hovoříme o tzv. *multimodální analýze sentimentu*. [27] V této práci se však pozornost soustředí výhradně na analýzu sentimentu v textu.

## 2.1 Techniky analýzy sentimentu

K analýze sentimentu lze přistupovat různými způsoby, které se liší technickým zázemím i mírou využití jazykových modelů, představených v sekci 1.2. Zpravidla se rozlišují metody slovníkové (lexicon-based), tradiční přístupy strojového učení, hluboké učení (neuronové sítě) a hybridní řešení. K modernějším technikám pak patří postupy založené na transformátorové architektuře (viz podsekcce 1.2.2.1). Podrobný přehled výhod a nevýhod těchto metod i jejich konkrétních možností nasazení lze nalézt v díle [28].

V této práci je pro analýzu sentimentu zvolen přístup využívající transformátorové modely, jak bylo podrobněji popsáno v sekci 1.2.2. Díky mechanismu pozornosti (attention) tyto modely lépe zachycují kontext a jemné jazykové nuance, a proto mnohdy překonávají klasické techniky analýzy sentimentu v přesnosti i rozsahu využití.

### 2.1.1 Slovníková metoda

Slovníkový přístup vychází z předpokladu, že některá slova bývají tradičně vnímána jako „pozitivní“ a jiná jako „negativní“. Tato slova se shromažďují v tzv. slovnících (lexikonech), podle nichž se následně určuje sentiment textu. Rozlišují se přitom dva základní postupy, korpusový a slovníkový, přičemž každý z nich má své specifické výhody i nevýhody. Detailnější přehled této metody lze nalézt v pracích [26, 28].

### 2.1.2 Tradiční přístupy strojového učení

Tradiční metody strojového učení vycházejí z rozdělení dat na trénovací a testovací sadu. Model se na trénovací sadě naučí rozlišovat sentiment (případně jiné charakteristiky) a poté se ověří na dosud neviděných datech z testovací sady. Příkladem takových postupů jsou Naivní Bayes (NB), Metoda podpůrných vektorů (SVM) nebo rozhodovací stromy [28]. Různé studie zkoumající nasazení těchto tradičních metod v analýze sentimentu jsou přehledně popsány v [29].

### 2.1.3 Hluboké učení

Kvůli nedostatečnému výpočetnímu výkonu nebylo možné hluboké učení (neuronové sítě) důkladně zkoumat. Avšak díky výraznému technologickému pokroku, zejména v oblasti grafických procesorů (GPU), získaly neuronové sítě na výkonnosti a možnosti využití. Hluboké učení využívá kaskádu více vrstev nelineárních zpracovatelských jednotek pro extrakci a transformaci příznaků.

Nižší vrstvy, které jsou blízko vstupním datům, se učí jednoduché rysy, zatímco vyšší vrstvy zachycují složitější vlastnosti, jež jsou odvozeny z příznaků získaných v nižších vrstvách. [30]

V této sekci budou představeny nejpoužívanější typy neuronových sítí v oblasti NLP, a to zejména konvoluční neuronové síť (*Convolutional Neural Networks, CNN*), rekurentní neuronové síť (*Recurrent Neural Networks, RNN*) a síť s dlouhou krátkodobou pamětí (*Long Short Term Memory networks, LSTM*). Jednotlivé typy neuronových sítí jsou zde popsány pouze stručně, zatímco podrobnější informace lze nalézt v díle [30]. Přehled studií zkoumajících využití těchto metod v praxi je uveden v [29].

### 2.1.3.1 Konvoluční neuronové síť

Konvoluční neuronová síť (CNN) je druh dopředné neuronové sítě, který byl původně vyvíjen pro úlohy spojené s počítačovým viděním. Její architektura čerpá inspiraci z lidské zrakové kůry (vizuální mechanismus přítomný v mozku živočichů). Tato zraková kůra obsahuje početné skupiny buněk citlivých na světlo v malých, vzájemně se překrývajících oblastech zorného pole, jež se označují jako receptivní pole. Tyto buňky fungují jako lokální filtry, které přenášejí informace o vstupních signálech dále. CNN se proto skládá z několika konvolučních vrstev, přičemž každá vrstva napodobuje funkci zmíněných buněk v zrakovém centru. [30]

### 2.1.3.2 Rekurentní neuronové síť

Rekurentní neuronová síť (RNN) [31] je typ neuronové sítě, ve které jsou spojení mezi neurony uspořádána do směřovaného cyklu. Na rozdíl od dopředných neuronových sítí dokáže díky své vnitřní „paměti“ zpracovávat i sekvenční data. Tato „paměť“ znamená, že pro každý prvek vstupní sekvence používá RNN stejné operace, přičemž výstup v každém kroku závisí na všech dosavadních výpočtech. Tímto způsobem si RNN „pamatuje“ kontext již zpracovaných prvků a uplatňuje jej při dalším výpočtu. [30]

### 2.1.3.3 Síť s dlouhou krátkodobou pamětí

Síť s dlouhou krátkodobou pamětí (LSTM) [32] představuje speciální druh RNN schopný pracovat s dlouhodobými závislostmi. Všechny RNN lze popsat jako řetězec opakujících se modulů, avšak u klasické RNN je tento modul zpravidla jednoduchý, zatímco LSTM využívá složitější vnitřní strukturu. Namísto jedné vrstvy neuronů obsahuje čtyři, které spolu vzájemně specifickým způsobem komunikují. [30]

#### 2.1.3.4 Další metody hlubokého učení v NLP

Neuronové sítě a hluboké učení nejsou omezeny pouze na výše popsané tři typy. Existuje řada dalších přístupů a využití, která zde budou zmíněna jen stručně, podrobnější informace lze nalézt v díle [30].

**Mechanismus pozornosti s RNN** Aby bylo možné lépe zachytit dlouhodobé závislosti, s nimiž měly RNN a LSTM obtíže, začal se využívat mechanismus pozornosti, inspirovaný vizuální pozorností u člověka. Prvně byl aplikován u strojového překladu v NLP [33]. Tyto sítě využívaly strukturu *encoder-decoder*, která byla již letmo představena v sekci 1.2.3 a později se stala hlavním stavebním kamenem transformátorové architektury. [30]

**Paměťové sítě** Paměťové sítě (*Memory networks, MemNN*) byly původně představeny v souvislosti s úlohou odpovídání na otázky [34]. Využívají několik inferenčních komponent v kombinaci s velkou dlouhodobou pamětí, přičemž tyto komponenty mohou být tvořeny neuronovými sítěmi. Na jejich základě vznikla vylepšená verze End-to-End paměťová síť (*End-to-End Memory network, MemN2N*) [35], která kombinuje mechanismus pozornosti a práci s dlouhodobou pamětí. [30]

**Rekurzivní neuronové sítě** Rekurzivní neuronové sítě (*Recursive neural networks, RecNN*) představují typ neuronových sítí, které se obvykle používají k práci s řízenými acyklickými grafy (typicky stromovými strukturami). Jedná se o obecnější variantu rekurentních neuronových sítí (RNN). [30]

#### 2.1.4 Hybridní metoda

Jak název napovídá, hybridní metoda propojuje více přístupů – například kombinuje charakteristiky slovníkových metod a strojového učení, aby využila silné stránky jednotlivých modelů a kompenzovala jejich slabiny [26].

## 2.2 Úrovně analýzy sentimentu

Výzkumníci obvykle zkoumají sentiment na třech základních úrovních podrobnosti: *dokumentové, větné a aspektové*. [28, 30]

### 2.2.1 Dokumentová úroveň

Klasifikace na dokumentové úrovni posuzuje celý text (např. recenzi produktu) jako jeden celek a určuje, jaký vyjadřuje postoj. Dokument se typicky zaměřuje na jednu entitu (např. „telefon“) a snaží se posoudit její celkový sentiment. [30]

Z pohledu analýzy sentimentu se v rámci této úrovně často pracuje jako s klasifikací dokumentu do tří tříd (pozitivní, negativní a neutrální). Tradičně

se využíval přístup *Bag of Words*, který s textem zachází jako s množinou slov bez ohledu na pořadí. O výsledném sentimentu rozhodovala frekvence nebo přítomnost určitých výrazů. Později se začaly prosazovat neuronové sítě, které využívají kódování slov do tzv. *dense vectors*, které dokáží zakódovat informace o sémantických a syntaktických vlastnostech. Podrobnější informace o těchto přístupech a přehled výzkumů zaměřených na dokumentovou úroveň sentimentu lze nalézt v [30].

### 2.2.2 Větná úroveň

Na této úrovni se sentiment klasifikuje pro jednotlivé věty v textu. Obvykle se nejprve rozhoduje, zda věta vůbec vyjadřuje názor či nikoliv (tzv. *subjectivity classification*), a teprve následně se stanovuje, jestli je vyjádřený sentiment pozitivní, negativní, nebo neutrální (tzv. *polarity classification*) [30].

Samotná predikce sentimentu pak probíhá podobně jako u dokumentové úrovně, ale vzhledem ke kratšímu rozsahu vět lze efektivněji využít i detailnější syntaktické a sémantické informace. Podrobnější informace o analýze sentimentu na větné úrovni a přehled souvisejících výzkumů jsou uvedeny v [30].

### 2.2.3 Aspektová úroveň

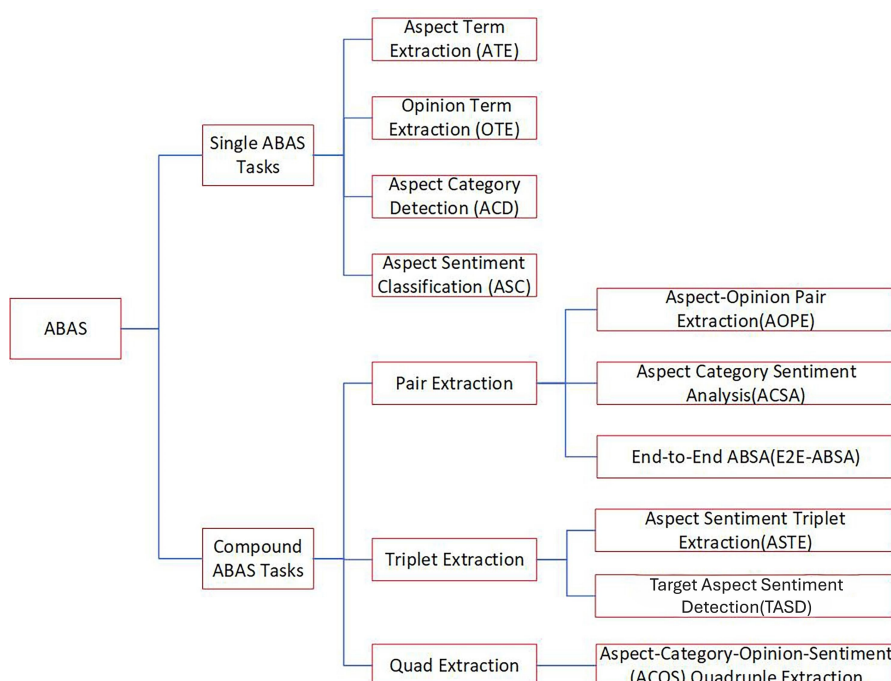
Aspektová úroveň představuje nejdetailnější formu analýzy sentimentu, protože nezkoumá pouze celkový sentiment textu, ale soustředí se na jednotlivé vlastnosti (*aspekty* nebo *cíle*) konkrétní entity. [30] Tato úroveň představuje hlavní zaměření této práce, neboť umožňuje nejpřesněji interpretovat recenze či analyzovat mediální texty, například s cílem zjistit, jak jsou v novinových článcích hodnoceny různé firmy nebo osobnosti. Podrobnější popis celé problematiky aspektové analýzy sentimentu bude v následující sekci 2.3.

## 2.3 Aspektová analýza sentimentu

Aspektová analýza sentimentu (ABSA, *Aspect-based sentiment analysis*) je detailní forma analýzy sentimentu, která v poslední době získává významnou pozornost ve výzkumu. Hlavní myšlenkou ABSA je, že emoce nebo názory obsažené v textu se nevztahují pouze na jednu entitu jako celek, ale konkrétně na její určité aspekty či vlastnosti [36]. Výzkum ABSA se obecně zaměřuje na identifikaci sentimentu z několika vzájemně propojených pohledů, zejména na úrovni aspektů, aspektových kategorií, názorových výrazů a polarity sentimentu [37]. [28]

Tato úroveň analýzy sentimentu představuje nejdetailnější stupeň, který lze dále rozčlenit do několika dílčích úloh, zaměřujících se na různé aspekty a jejich vlastnosti. V následujících podsekcích budou nejprve představeny jednotlivé dílčí úlohy aspektové analýzy sentimentu (*Single ABSA Tasks*), a poté budou popsány složitější úlohy (*Compound ABSA Tasks*), vznikající kombinací těchto





■ **Obrázek 2.1** Jednotlivé dílčí úlohy aspektové analýzy sentimentu, upravený obrázek z [28]

dílčích úloh, které již poskytují komplexní informace o sentimentu vzhledem k jednotlivým aspektům (viz. Obrázek 2.1).

### 2.3.1 Základní dílčí úlohy ABSA

Aspektová analýza sentimentu zahrnuje čtyři základní dílčí úlohy (tzv. *Single ABSA Tasks*). Konkrétně se jedná o extrakci aspektů (*Aspect Term Extraction, ATE*), extrakci názorových výrazů (*Opinion Term Extraction, OTE*), detekci aspektových kategorií (*Aspect Category Detection, ACD*) a klasifikaci sentimentu aspektů (*Aspect Sentiment Classification, ASC*). [28]

#### ■ Extrakce aspektů (ATE)

Cílem této úlohy je identifikace všech aspektů (neboli entit) zmíněných v textu, vůči kterým se následně provádí analýza sentimentu. Jedná se o klíčový krok většiny komplexnějších ABSA úloh, protože definuje hlavní objekty další analýzy. [38]

#### ■ Extrakce názorových výrazů (OTE)

Tato úloha spočívá ve vyhledávání slov či frází, které vyjadřují názory nebo postoje, aniž by nutně specifikovaly konkrétní aspekty. Úkolem je tedy získat ze samotného textu výrazy, které se následně přiřazují k relevantním aspektům. [38]



### ■ Detekce aspektových kategorií (ACD)

Tato úloha se věnuje určování obecných kategorií aspektů. Existují zde dva přístupy. První spočívá v kategorizaci aspektů do širších skupin (např. ve větě „Salát byl skvělý, ale náš číšník byl velmi nepříjemný“ by výsledkem byly kategorie „jídlo“ a „obsluha“ [28]). Druhý přístup využívá kategorie k popisu vlastností jednotlivých aspektů (např. věta „Pokoje jsou skvělé, jsou velmi levné“ by vedla ke kategoriím „obecná“ a „cena“ [39]).

### ■ Klasifikace sentimentu aspektů (ASC)

Tato úloha je přímo zaměřena na určení sentimentu konkrétních aspektů zmíněných v textu. Jde tedy o samotnou analýzu sentimentu, kdy cílem je přiřadit specifikovaným aspektům správnou polaritu sentimentu (pozitivní, negativní či neutrální). Právě tato úloha představuje hlavní zaměření této bakalářské práce. [38]

## 2.3.2 Složené úlohy ABSA

Jak již bylo zmíněno, složené úlohy ABSA (tzv. *Compound ABSA Tasks*) spojují několik dílčích úloh dohromady, aby bylo možné lépe zachytit a popsat sentiment obsažený v textech. Hlavním cílem těchto úloh je extrahovat a predikovat více sentimentových prvků současně. Mezi typy složených úloh patří extrakce dvojic, trojic a čtveřic sentimentových prvků. [28]. Jelikož jsou tyto úlohy již nad rámec této práce, budou popsány pouze stručně, podrobnější popis je dostupný v citacích u jednotlivých úloh.

### 2.3.2.1 Extrakce dvojic

Mezi úlohy založené na extrakci dvojic patří extrakce aspekt-názor dvojic (*Aspect-Opinion Pair Extraction, AOPE*), analýza sentimentu aspektových kategorií (*Aspect Category Sentiment Analysis, ACSA*) a end-to-end aspektová analýza sentimentu (*End-to-End ABSA, E2E-ABSA*) [28], někdy označovaná jako jednotná aspektová analýza sentimentu (*Unified ABSA, UABSA*) [40]. Tyto úlohy poskytují ucelenější pohled na sentiment tím, že spojují jednotlivé prvky analýzy (např. aspekt a odpovídající názor) nebo zároveň určují sentiment vůči daným aspektům.

### ■ Extrakce aspekt-názor dvojic (AOPE)

Tato úloha spočívá v extrakci párů tvořených aspektem a k němu náležícím názorovým výrazem. Cílem je identifikovat vazbu mezi konkrétními aspekty a slovy, která je popisují. Například z věty: „Salát byl skvělý a číšník byl také velmi nápomocný“ se získají dvojice: „[salát, skvělý]“ a „[číšník, nápomocný]“. [40]

### ■ Analýza sentimentu aspektových kategorií (ACSA)

Při této úloze se analyzuje sentiment v rámci předem definovaných aspektových kategorií, a tím se určuje celková polarita dané kategorie. Například

z věty: „Pití bylo skvělé a jídlo bylo velmi chutné“ by se získaly dvojice: „[občerstvení, pozitivní]“ a „[občerstvení, pozitivní]“, což dohromady vypovídá o pozitivním sentimentu pro kategorii „občerstvení“. [41]

#### ■ End-to-End aspektová analýza sentimentu (E2E-ABSA)

Tato úloha v jednom kroku spojuje extrakci aspektových termínů a jejich klasifikaci sentimentu, tedy zároveň identifikuje aspekty i jejich náladu (polaritu). Například z věty: „Salát byl skvělý a číšník byl také velmi nápomocný“ se získají dvojice: „[salát, pozitivní]“ a „[číšník, pozitivní]“. [40]

### 2.3.2.2 Extrakce trojic

Mezi úlohy založené na extrakci trojic patří extrakce aspekt-názor-sentiment trojic (*Aspect Sentiment Triplet Extraction, ASTE*) a detekce aspekt-kategorie-sentiment (*Target Aspect Sentiment Detection, TASD*). [28] Tyto úlohy přinášejí podrobnější popis sentimentu tím, že identifikují nejen aspekty, ale i jejich názory či vlastnosti a sentiment současně.

#### ■ Extrakce aspekt-názor-sentiment trojic (ASTE)

Cílem této úlohy je identifikovat trojici tvořenou aspektem, k němu příslušným názorovým výrazem a polaritou sentimentu. Poskytuje detailní a přesný vhled do hodnocení, protože zohledňuje, *co* je hodnoceno (aspekt), *jaký* je k němu sentiment a případně *proč* (na základě čeho). Například z věty: „Salát byl skvělý a číšník byl také velmi nápomocný“ lze získat trojice: „[salát, skvělý, pozitivní]“ a „[číšník, nápomocný, pozitivní]“. [28, 42]

#### ■ Cílená aspektová analýza sentimentu (TASD)

Tato úloha spočívá v detekci kategorií aspektů, konkrétních aspektů a jejich sentimentové polaritě. Hlavním cílem je určit, do jaké kategorie daný aspekt spadá a jakou má polaritu. Například z věty „I přesto, že pizza byla vážně skvělá, cena byla velmi vysoká“ lze odvodit trojice: „[pizza, jídlo#kvalita, pozitivní]“ a „[pizza, jídlo#cena, negativní]“, čímž se vyjádří, že pizza je sice kvalitní, avšak poměrně drahá. [43, 44]

### 2.3.2.3 Extrakce čtveřic

Nejkomplexnější úlohou v ABSA je extrakce čtveřic, konkrétně jde o extrakci čtveřic aspekt-kategorie-názor-sentiment (*Aspect-Category-Opinion-Sentiment Quadruple Extraction, ACOS*) [28]. Tato úloha integruje všechny základní prvky aspektové analýzy sentimentu a umožňuje nejdetailnější pohled na postoje vyjádřené v textu.

#### ■ Extrakce čtveřic aspekt-kategorie-názor-sentiment (ACOS)

Tato úloha představuje nejkomplexnější formu extrakce, při které se získávají kompletní čtveřice obsahující konkrétní aspekt, jeho kategorii, odpovídající názor a sentiment. Jako příklad lze uvést větu: „Vypadá pěkně a jeho

konstrukce je velice bytelná.“ Z této věty je možné získat čtveřice: „[povrch, design, bytelná, pozitivní]“ a „[null, design, pěkně, pozitivní]“. V některých případech nemusí být všechny složky čtveřice ve větě explicitně vyjádřeny, což představuje další výzvu této úlohy. [28, 45]

### 2.3.3 Využití transformátorových modelů v ABSA

V sekci 2.1 byly představeny různé metody analýzy sentimentu. Stále více studií se však v současné době zaměřuje na využití transformátorové architektury [16], například modelu BERT [9] a jeho odvozených variant. Tato část proto popisuje práce, které aplikují BERT na jednotlivé metody ABSA, srovnávají jej s dalšími modely a zkoumají jeho vhodnost v různých doménách i jazycích. Také v této práci budou použity vybrané varianty modelů BERT, a to jak pro různé jazyky, tak pro různé okruhy dat.

#### 2.3.3.1 Datasets

Kvalitní a dostatečně rozsáhlá trénovací data představují klíčový předpoklad úspěchu při trénování modelů. V kontextu analýzy sentimentu, a především u složitějších úloh ABSA, ovšem není dostupných datasetů mnoho. Proto některé studie usilují o tvorbu nových datových sad, jež obsahují větší množství domén, více sentimentů v jedné větě či rozsáhlejší objem dat. Na těchto nových korpusech následně testují výkonnost modelů typu BERT, aby ověřily jejich schopnost zachytit jemné nuance aspektové analýzy.

##### ■ Dataset OATS

V datové sadě OATS (*Opinion Aspect Target Sentiment*) [39] je cílem připravit objemnější datasety a rozšířit dosavadní spektrum domén (často omezené na recenze notebooků nebo restaurací [46]) o další oblasti, jako jsou potraviny (Amazon Food Reviews), online kurzy (Coursera) a hotely (TripAdvisor). Tento dataset je připraven pro několik složených ABSA úloh (E2E-ABSA, ASTE, TASD a ACOS).

##### ■ Dataset MAMS

Cílem datové sady MAMS (*Multi-Aspect Multi-Sentiment*) [47] je zajistit, aby každá věta obsahovala alespoň dva hodnocené aspekty. Data pocházejí z kolekce CitySearch New York a jsou uzpůsobena primárně pro úlohy E2E-ABSA a ACSA.

#### 2.3.3.2 Složené úlohy ABSA

Pro jednotlivé složené úlohy ABSA existuje řada studií, které se zaměřují na hlubší pochopení daných problémů či na formulaci nových variant těchto úloh. V této sekci jsou stručně představeny vybrané práce, jež nabízejí nové pohledy a přístupy.

### ■ ACSA

- Jednou z hlavních výzev ACSA je skutečnost, že k jedné kategorii se v jedné větě může vztahovat více názorů. Tento problém řeší například autoři studie [41], kteří nejprve vyhledají v textu konkrétní instance aspektových kategorií, přiřadí jim sentiment a následně určí výslednou polaritu pro danou kategorii.
- V práci [48] autoři navrhnou model KASA, který tuto úlohu pojímá jako klasifikaci dvojice vět (sentence-pair classification) namísto klasifikace jedné věty (single-sentence classification), čímž ji činí jednodušší k řešení.

### ■ E2E-ABSA

- V [49] je popsáno využití modelu BERT a jeho upravených verzí pro End-to-End ABSA. Ve srovnání se staršími modely (např. LSTM) dosáhl BERT výrazně lepších výsledků. Navíc autoři sjednocují porovnání modelů tím, že důsledně využívají samostatný (hold-out) vývojový dataset pro ladění modelu, což dřívější studie často přehlížely. Díky tomu jejich řešení slouží jako BERT-based benchmark pro E2E-ABSA.
- Autoři studie [50] se naopak soustředí na vývoj takového modelu pro analýzu sentimentu, který lépe obstojí v různých doménách, a snaží se tak vyřešit problém omezení modelů na konkrétní doménu, jež vyplývá z použití doménově specifických datasetů pro fine-tuning.

### ■ ASTE

- Práce [51] převádí reprezentaci trojic do „relační oblasti“ (relation region) v 2D tabulce a pojímá úlohu ASTE jako detekci a klasifikaci těchto oblastí.
- Ve studii [52] je představen „přístup bez značkování“ (tagging-free), který lépe zachycuje vzájemnou závislost mezi komponentami trojice. Autoři využívají architekturu typu encoder-decoder v kombinaci s modelem BERT.

### ■ TASD

- Ve studii [53] autoři rozšiřují BERT o dodatečný kontext a sledují dopady této úpravy na výkonnost modelu. Zároveň poukazují na to, že rozšíření předtrénovaných modelů typu self-attention o kontextové závislosti dokáže významně zlepšit výsledky v kontextově orientovaných jazykových úlohách.
- V [54] se ověřuje, nakolik se zlepší přesnost predikce, pokud je *TASD* pojato jako klasifikace dvojice vět (sentence-pair classification), například formou otázky a odpovědi. Tento přístup je použit i v této práci, kde bude porovnán s klasifikací jedné věty (single-sentence classification).

- Někdy aspekt (target) není ve větě explicitně zmíněn, nýbrž pouze naznačen. Příklad: „Jídlo přišlo po 20 minutách a bylo studené.“ – je zde explicitní trojice s aspektem „jídlo“, avšak lze rovněž odvodit negativní hodnocení „obsluhy“, neboť jídlo dorazilo pozdě. Tím se zabývá studie [43].
- V [44] se výzkumníci snaží řešit *TASD* prostřednictvím „systému využívajícího prompty“ (prompt-based system), kde je generována otázka a k ní odpověď. Autoři rovněž navrhují vlastní řešení, jak tento přístup aplikovat v praxi.

#### ■ ACOS

- Ve studii [45] autoři popisují tuto komplexní úlohu, která propojuje všechny dílčí ABSA úlohy, a zároveň vytvářejí datasety Restaurant-ACOS a Laptop-ACOS. Své řešení testují na čtyřech různých systémech, jejichž podrobný popis je uveden v originálním díle.
- V [55] je navržena metoda pro extrakci čtveřic, jež zohledňuje vzdálenost mezi aspekty a názorovými výrazy. Autoři následně porovnávají své výsledky na datasetech Restaurant-ACOS a Laptop-ACOS s jinými přístupy, přičemž dosahují lepších výkonnostních ukazatelů.

## 2.4 Využití analýzy sentimentu

Analýza sentimentu nachází uplatnění v celé řadě oblastí. Díky rozmachu internetu a snadné dostupnosti pro uživatele sdílet své názory a pocity se význam tohoto přístupu neustále zvyšuje. Firmy a organizace mohou na základě recenzí a komentářů efektivně upravovat své produkty a služby, a tím zvyšovat zisk. V oblasti zdravotnictví lze analýzu sentimentu využít například ke sledování potenciálních sebevražedných myšlenek prostřednictvím detekce negativních emocí ve veřejně publikovaných příspěvcích. [27] V následujících podsekcích jsou stručně představeny hlavní oblasti, kde se analýza sentimentu uplatňuje, a uvedeny studie, které se danou tematikou zabývají.

### 2.4.1 Obchodní oblast

Analýza sentimentu představuje v obchodní sféře jeden z nejvýznamnějších nástrojů pro zlepšování produktů a služeb. Díky schopnosti zpracovat velké množství recenzí lze účinně odhalovat nedostatky a identifikovat příležitosti ke zkvalitnění nabídky. Některé studie [56, 57] podrobně popisují proces extrakce sentimentu a jeho praktické využití v obchodním prostředí. Další práce se zaměřují na analýzu dat získaných z Amazonu [58], z čínských online recenzí [59], z Facebooku [60] či ze služby Zomato [61]. Souhrnně se těmito tématy zabývají také publikace [27, 28, 62].

Analýza sentimentu nemusí sloužit pouze velkým firmám a organizacím, ale nabízí také přínos pro jednotlivce zájímající se o finanční trhy, akcie či kryptoměny. Řada studií [63, 64, 65] se věnuje využití sentimentu při odhadech vývoje cen akcií a zkoumá, zda pozitivní či negativní nálada může signalizovat budoucí růst nebo pokles. Podobně existují i práce zaměřené na kryptoměny [66, 67], které zkoumají, do jaké míry lze sentiment využít pro predikci cenového vývoje. Přehledně se tímto tématem zabývají též díla [27, 28, 62].

Analýza sentimentu se nemusí zaměřovat pouze na recenze produktů a služeb, ale lze ji využít i k mapování obecných postojů k aktuálním událostem. Řada studií [68, 69, 70, 71, 72] například zkoumá reakce uživatelů sociálních sítí (zejména X, původně Twitter) na pandemii COVID-19. Další výzkumy se věnují jiným tématům, jako je rusko-ukrajinská válka [73] či postoje k elektromobilům v kontextu rostoucích cen pohonných hmot [74]. Přehledný souhrn těchto výzkumů nabízí také články [27, 28].

### 2.4.2 Doporučující systémy

Analýza sentimentu nachází uplatnění i v doporučujících systémech, ať už jde o hotely či filmy. Typickým přístupem je práce s recenzemi konkrétního uživatele, na jejichž základě lze určit, co by mu mohlo být doporučeno. Různé metody a postupy pro návrh doporučujících systémů popisují studie [75, 76, 77, 78, 79], jejichž souhrn je uveden v [62].

### 2.4.3 Zdravotnická oblast

V předchozích oblastech byla analýza sentimentu popsána především ve vztahu k ekonomickým přínosům, ať už pro firmy nebo pro jednotlivce. Tato metoda však nachází uplatnění i při zlepšování zdravotního stavu populace. Lze ji například využít ke sledování názorů veřejnosti na lékaře, léčiva a zdravotní péči [80, 81, 82], nebo k monitorování ohrožených skupin, jako jsou senioři [83], případně k detekci sebevražedných myšlenek a depresí na základě příspěvků na sociálních sítích [84, 85, 86]. Dokonce existují studie, které se zaměřují na predikci výskytu rakoviny tlustého střeva a konečníku [87]. Podrobnější popis uvedených aplikací nabízí např. díla [27, 28, 62].

Analýza sentimentu se přitom nemusí týkat pouze fyzického zdraví. Lze ji použít i pro prevenci vulgárních či nenávistných projevů na internetu a tím bránit psychické zdraví online uživatelů. Negativní a urážlivé výrazy mohou u uživatelů vyvolávat stres nebo přispívat k rozvoji depresí. Řada studií se proto zabývá detekcí nenávistných projevů (hate speech) například na X (Twitter) [88, 89], Facebooku [90] či analýzou kyberšikany na Instagramu [91]. Souhrnné informace o těchto využitích lze nalézt v [27].

## 2.5 Otevřené problémy analýzy sentimentu

Ačkoli se koncept analýzy sentimentu může na první pohled zdát jednoduchý (určitý názor je buď pozitivní, negativní či neutrální), v praxi se často objevují situace, kdy tatáž slova mohou v různém kontextu nést kladný i záporný význam. Člověk tyto nuance rozeznává intuitivně, avšak pro počítačové zpracování představují značnou výzvu. Následující paragrafy nastíní některé z hlavních problémů a otevřených otázek v oblasti analýzy sentimentu. Podrobnější rozbor a další problémy lze nalézt v [27, 28, 62].

**Sarkasmus** Sarkasmus se v běžném životě vyskytuje poměrně často, někteří lidé ho používají prakticky denně. Pro klasickou analýzu textu však představuje problém, protože věta, která vypadá na první pohled pozitivně, může být ve skutečnosti míněna ironicky. Různé studie uvedené v [62] se proto zaměřují na rozpoznávání a správnou interpretaci sarkasmu.

**Vícejazyčná analýza sentimentu** Většina současných výzkumů se soustředí převážně na angličtinu, jelikož pro ni existuje relativně velké množství dat a dalších jazykových zdrojů. U jazyků s nedostatečnou podporou (tzv. *low-resource languages*) je však problematické získat dostatečné datasety pro předtrénování modelů, natož pro samotnou analýzu sentimentu [62]. Komplikací bývají i texty, v nichž se mísí více jazyků (tzv. *code-switching*), jelikož pro tuto oblast dosud chybějí rozsáhlejší zdroje dat. Podrobnější rozbor těchto obtíží je k dispozici v [27, 28, 62].

**Problém domén** Datasety pro trénování analýzy sentimentu obvykle pocházejí pouze z jedné oblasti, například z recenzí notebooků či restaurací [46]. Modely se tak naučí rozpoznávat sentiment především v dané doméně, což je nevýhodou při nasazení do jiného kontextu. Některé výrazy, které mají v recenzích notebooků pozitivní význam, se totiž v jiných oblastech nemusí projevovat stejně. [27, 28, 62]

## Český jazyk a analýza sentimentu

Výzkum zpracování přirozeného jazyka se soustředí především na nejčastěji používané jazyky, zejména angličtinu a čínštinu. Tyto jazyky disponují rozsáhlými korpusy, dobře zdokumentovanými benchmarky a nové modely jsou na nich zpravidla validovány jako první (například *ModernBERT* [24], který je v době psaní dostupný pouze pro angličtinu).

U méně rozšířených jazyků, k nimž patří čeština, je situace odlišná. Omezená dostupnost dat, menší komunita a morfologická složitost vedou k nižší míře experimentálního ověřování i praktických aplikací. Tato kapitola proto shrnuje dosavadní práci v oblasti analýzy sentimentu pro český jazyk, popisuje přínos publikovaných studií a diskutuje jejich vliv na další rozvoj metod.

### 3.1 Předešlé práce na analýzu sentimentu v češtině

Tato sekce se zaměřuje na jednotlivé práce a studie, které se věnují různým aspektům analýzy sentimentu v češtině. Pokrývá témata jako definice sentimentu v českém jazyce, porovnání různých modelů na českých datasetech, představování nových datasetů, a dokonce i vývoj jazykových modelů specializovaných pro češtinu.

#### 3.1.1 Problematika analýzy sentimentu v češtině

V knize [92] autorka detailně zkoumá problematiku sentimentu v českém jazyce. Popisuje různé jazykové roviny, které byly představeny v sekci 1.1.1.1, a vysvětluje jejich chování v českém jazyce, přičemž uvádí množství konkrétních příkladů. Také se zaměřuje na emocionální struktury (Formální reprezentace emocionálních struktur), přičemž popisuje jejich analýzu a roli v rámci sentimentu.



Autorka se dále věnuje různým podúkolům. Nejprve zkoumá analýzu sentimentu na větné úrovni v českém jazyce. V této sekci jsou porovnány tři klasifikátory, které jsou testovány na různých datasetech z českých webových stránek. [92]

Následně se zaměřuje na aspektovou analýzu sentimentu, která využívá dataset SemEval-2014 [46], stejně jako v této práci. Avšak „targety“ jsou hledány pomocí již známých targetů z trénovacích dat, což znamená, že tento model není dostatečně robustní a nebude fungovat v jiných doménách. Tuto vytvořenou pipeline pak adaptovala i pro češtinu. [92]

V poslední části autorka popisuje řešení problému TASD v rámci SemEval-2016 Task 5 [93]. Pro tento úkol byly použity metody hlubokého učení, konkrétně LSTM. Tento přístup vedl k vítězství ve dvou kategoriích (jazycích), ruštině a turečtině. Avšak výsledky nebyly tak uspokojivé v jiných jazycích. [92]

### 3.1.2 Porovnání různých modelů

V díle [94] si autoři kladli za cíl provést přehled a porovnání různých tradičních přístupů strojového učení. Použili dva datasety, jeden z Facebooku a druhý z Mall.cz, přičemž se zaměřili pouze na analýzu sentimentu na větné úrovni. Testovali různé modely, jako jsou SVM, logistická regrese (Maximum Entropy) a Naivní Bayes. Dále vyzkoušeli i náhodné lesy a vícevrstvé neuronové sítě. Nejlepší výsledky dosáhl model SVM na datasetu Mall.cz a model logistické regrese na datasetu Facebook.

### 3.1.3 Představení nových technik

Zajímavý přístup k řešení ACSA předkládají autoři v díle [95]. Spojují dvě úlohy z NLP: aspektovou analýzu sentimentu (ABSA) a sémantické značkování rolí (*Semantic Role Labeling, SRL*). Využívají model ELECTRA [96], který je výpočetně méně náročný než modely založené na BERTu. Představují několik způsobů propojení ABSA encoderu a SRL encoderu. V češtině dosáhli zlepšení v obou podúlohách (extrakce kategorií a určení jejich sentimentu), zatímco v angličtině se výsledky nezvýšily – tam nadále dominují rozsáhlejší modely typu BERT.

### 3.1.4 Datasety v češtině

Jedním z hlavních úskalí analýzy sentimentu v češtině je omezená dostupnost kvalitních datových sad. Na rozdíl od angličtiny, pro niž existuje rozsáhlé množství datasetů, zůstává čeština podstatně hůře pokryta. Řada studií se proto zaměřuje na tvorbu nových datasetů, jež mají sloužit jako základ pro další rozvoj a zlepšování modelů v českém jazyce. V této části jsou představeny datové sady určené jak pro sentiment na větné úrovni, tak pro aspektovou analýzu sentimentu.

#### 3.1.4.1 Analýza sentimentu na větné úrovni

Tvorbou větně anotovaného datasetu se zabývají autoři v článku [97]. Autoři shromáždili příspěvky z devíti českých facebookových stránek a manuálně je označili sentimentem, aby dataset posloužil pro trénink různých modelů. Porovnali výkon SVM a logistické regrese (v článku označované jako *Maximum Entropy*) na vlastním korpusu i na filmových recenzích z ČSFD a produktových recenzích z Mall.cz. Pro každý model otestovali více sad vstupních příznaků a sledovali jejich vliv na přesnost. Výsledky se pohybovaly na podobné úrovni jako ve studii [94], která s tímto korpusem experimentovala s širším spektrem modelů.

#### 3.1.4.2 Aspektová analýza sentimentu

Ve studii [98] autoři představili první český dataset připravený pro ABSA. Korpus vychází z recenzí restaurací, podobně jako dataset SemEval-2014 [46], a je anotován pro úlohy ACSA i E2E-ABSA. V této práci bude použit jako jeden z datasetů, na kterých budou natrénovány všechny vybrané modely a následně budou porovnány výsledky mezi češtinou a angličtinou.

Na tento korpus navázala práce [99], jejímž cílem bylo vytvořit dataset pro náročnější úlohy ABSA, konkrétně TASD. Nový dataset byl otestován také na jednodušších úlohách, například ATE, ACD a E2E-ABSA. Autoři se inspirovali strukturou datasetu SemEval-2016 [93], který je připraven ve stejném formátu pro více jazyků.

### 3.1.5 České jazykové modely

Většina nových architektur jazykových modelů je primárně validována na angličtině; vícejazyčné varianty sice později pokrývají až 100 jazyků, avšak jejich kapacita se mezi jazyky dělí nerovnoměrně. Specializované modely trénované čistě na českých datech proto často podávají stabilnější a přesnější výsledky. Následující studie se proto zaměřují na adaptaci vybraných architektur přímo pro češtinu. Všechny zde popsané modely (a další použité v experimentech) jsou podrobněji představené v sekci 4.3.

#### 3.1.5.1 Český BERT – Czert

Autoři studie [100] ze Západočeské univerzity vytvořili jazykový model Czert, který je jako první jazykový model natrénovaný výhradně na českých datech. Czert vznikl ve dvou variantách: Czert-B založený na architektuře BERT [9] a Czert-A vycházející z modelu ALBERT [101]. V této práci je použita varianta Czert-B, jejíž podrobnější popis lze nalézt v sekci 4.3.2.2.

### 3.1.5.2 Český RoBERTa – RobeCzech

Na Czert navázal model RobeCzech, který využívá architekturu RoBERTa a je rovněž předtrénován na českých datech. Autoři jej otestovali na pěti úlohách včetně analýzy sentimentu a ve všech případech překonal model Czert. [102] Podrobnější popis tohoto modelu lze nalézt v sekci 4.3.2.3.

### 3.1.5.3 FERNET – BERT i RoBERTa

Západočeská univerzita dále publikovala dvojici modelů FERNET (Flexible Embedding Representation NETwork) [103]. První varianta vychází z BERT-u, druhá z RoBERTa; obě rozšiřují dostupné české modely a umožňují širší srovnání na domácích datech. Podrobný popis přináší sekce 4.3.2.4.

## 3.2 Shrnutí dosavadního výzkumu

Jak ukazuje tento přehled, existuje řada studií zaměřujících se na NLP a analýzu sentimentu v českém jazyce. Tato díla přinášejí užitečné poznatky, ať už v oblasti analýzy sentimentu na větné úrovni, nebo v analýze sentimentu na úrovni aspektů. Většina těchto studií, která porovnává různé modely na různých datasetech, využívá tradiční metody strojového učení, jako jsou SVM, Naivní Bayes nebo logistická regrese. Tyto metody však, ve srovnání s moderními metodami hlubokého učení, konkrétně s transformátorovými modely, vykazují omezenější výkonnost.

Na druhou stranu, novější výzkumy, které se soustředí na tvorbu datasetů a vývoj českých jazykových modelů, se snaží připravit českou komunitu na použití novějších přístupů. Tyto studie nejen že vytvářejí potřebné korpusy pro trénování a testování modelů, ale také otevírají cestu k využívání transformátorových modelů, které jsou schopny efektivněji pracovat s jazykovými nuancemi a hledat ideální řešení pro různé úlohy ABSA. Tento přístup znamená výrazný krok vpřed, protože umožňuje dosáhnout vyšší přesnosti a lepších výsledků na složitějších úlohách analýzy sentimentu.

Tato práce navazuje na poznatky těchto studií a vychází z dosud zpracovaných datasetů pro českou aspektovou analýzu sentimentu. Dále bude využívat český jazykové modely a porovnávat jejich výkon mezi sebou a i s jinými modely, čímž bude přispívat k rozvoji analýzy sentimentu v češtině a hledání ideálních metod pro konkrétní úkoly ABSA.

# Metodologie a postup trénování

V předchozích kapitolách byly představeny klíčové poznatky z oblasti zpracování přirozeného jazyka, analýzy sentimentu a aspektové analýzy sentimentu. Tato kapitola se soustředí na všechny potřebné kroky nutné k trénování vybraných modelů, a proto zde budou popsány použité datasety, vybrané metody a modely. Dále je detailně vysvětlen postup trénování a představeny vyhodnocovací techniky spolu s praktickými testy modelů.

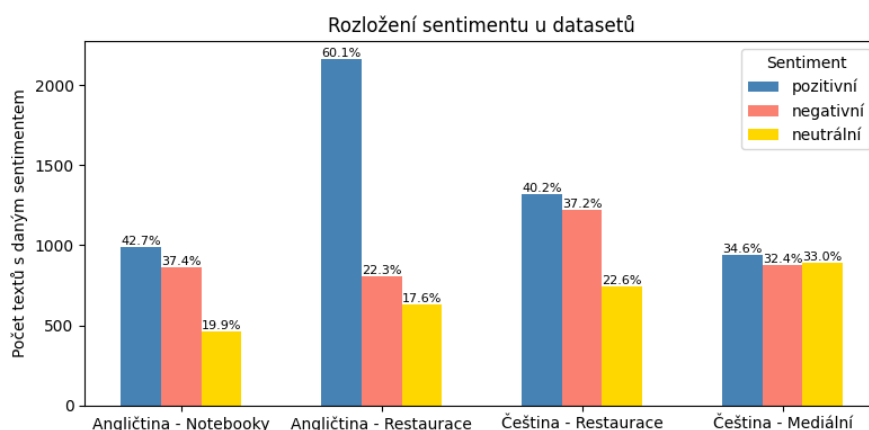
## 4.1 Datasety

Správný výběr dat představuje klíčový faktor pro úspěšné trénování jakéhokoliv modelu, od lineární regrese až po komplexní jazykové modely. Z tohoto důvodu jsou v této práci použity datové sady, které již v angličtině zaznamenaly výraznou pozornost výzkumné komunity. Mezi ně patří dataset SemEval-2014 [46], jenž patří k nejčastěji využívaným zdrojům pro ABSA. Pro české prostředí je analogicky použit obdobný dataset recenzí restaurací vytvořený na Západočeské univerzitě [98]. Nakonec budou použita také data z archivu Newton Media, která umožní otestovat modely v oblasti mediálních textů.

Každá datová sada je dále popsána samostatně – nejprve je představen její původ a způsob anotace, následně jsou shrnuty základní statistiky (počet záznamů, průměrná délka textu, distribuce polarit). Součástí popisu je rovněž srovnání různých tokenizátorů: u každého korpusu jsou uvedeny maximální délky tokenizovaných sekvencí, aby byl zřejmý rozdíl mezi tokenizátory trénovanými přímo na jazyku daného datasetu a těmi, které jsou optimalizovány pro jinou jazykovou rodinu. Podle maximální délky je pak vybrané omezení při trénování jednotlivých modelů. Tokenizátory a odpovídající jazykové modely jsou podrobně charakterizovány v sekci 4.3.

### 4.1.1 Příprava datasetů

Všechny datasety použité v této práci byly upraveny podle jednotného postupu. Nejprve byly vybrány pouze tři klíčové příznaky: „text“, „aspekt“ a „label“. Dataset byl následně očištěn od prázdných hodnot, čímž byla zajištěna kvalita dat pro následné zpracování. Tato práce se zaměřuje na tři hodnoty příznaku „label“, konkrétně na *pozitivní*, *negativní* a *neutrální*. Rozdělení jednotlivých datasetů podle „label“ je zobrazeno v grafu 4.1, který ukazuje i procentuální podíl jednotlivých hodnot „label“ v každém z datasetů.



**Obrázek 4.1** Rozdělení sentimentu v datasetech, vlastní práce

Pro správné použití datasetů při trénování modelů bylo nejprve nutné každý korpus rozdělit na tři části: *trénovací*, *validační* a *testovací*. Ve všech případech byl zachován jednotný poměr – 70 % dat bylo použito pro trénování, 20 % pro validaci a 10 % pro závěrečné testování. Předběžné dělení dat zajistí, že všechna následná trénování modelů využijí stejnou strukturu korpusu, a tím umožní objektivně porovnávat dosažené výsledky, protože trénování i vyhodnocení každého modelu proběhne nad totožnými, předem definovanými množinami.

### 4.1.2 Dataset SemEval-2014

Tato datová sada patřila k prvním, jež se soustředily na aspektovou analýzu sentimentu (ABSA). Do té doby se většina dostupných souborů pro analýzu sentimentu zaměřovala spíše na celkovou náladu či postoj v úrovni celé věty nebo dokumentu. Proto se tým za tímto projektem rozhodl připravit dvě samostatné kolekce recenzí – jednu zaměřenou na notebooky a druhou na restaurace. [46]

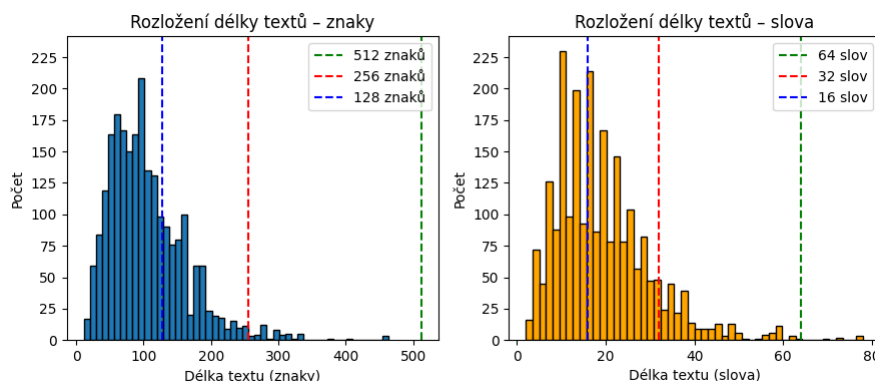
Oba datasety byly anotovány dvěma anotátory: jeden byl studentem magisterského studia a druhý odborníkem na lingvistiku. V případě neshod mezi těmito anotátory zasáhl třetí, expertní anotátor. Nejčastější neshody spadaly

do jednoho z následujících tří typů: *Nejasnost polarity* (Polarity ambiguity), *Hranice víceslovných aspektů* (Multi-word aspect term boundaries), a *termín aspektu versus odkaz na cílovou entitu* (Aspect term vs. reference to target entity). Podrobnější popis těchto typů a vytváření celého datasetu je uveden v jejich práci [46].

#### 4.1.2.1 Recenze notebooků

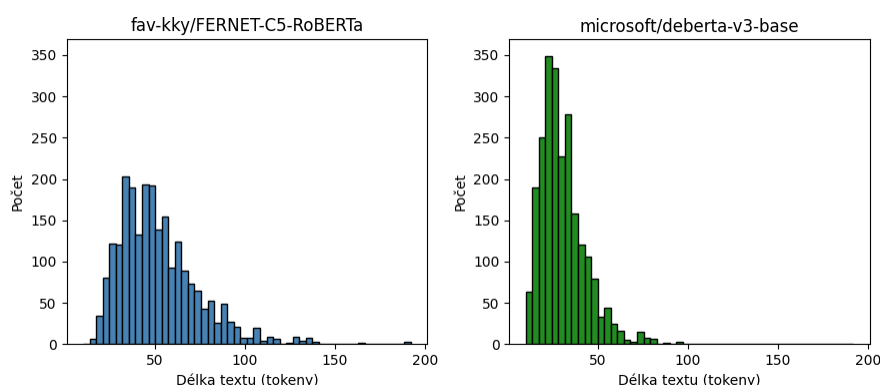
Jak již bylo zmíněno, tento dataset byl představen v díle [46]. Samotná data však byla získána z portálu Hugging Face, kde je připravil Tom Aarsen a jsou dostupná z [104].

V datasetu recenzí notebooků je celkem 2313 vzorků, rozdělených na 1619 trénovacích, 462 validačních a 232 testovacích. Tento dataset obsahuje 1456 unikátních textů a 1031 unikátních aspektů. Rozdělení sentimentu v tomto datasetu je znázorněno v grafu 4.1. Průměrná délka textu činí 104,8 znaků, což odpovídá přibližně 19,3 slovům. Nejdelší text má 464 znaků (nebo 78 slov). Celkové rozložení délky textu je zobrazeno v grafu 4.2.



■ **Obrázek 4.2** Rozdělení délky textů v datasetu Recenze notebooků v angličtině, vlastní práce

Mezi nejčastější aspekty v tomto datasetu patří: *screen* (obrazovka), *use* (použití) a *price* (cena). Pro tento dataset bylo testováno použití různých tokenizátorů, aby se zjistila ideální délka vstupních sekvencí pro modely. Maximální délky tokenizovaných sekvencí se pohybovaly v rozmezí 97 až 192 tokenů. Tento velký rozptyl je způsoben tím, že na anglický dataset byly použity i tokenizátory, které byly trénovány pouze na češtině, což vedlo k rozdílným výsledkům. Největší rozdíl v délkách tokenizovaných textů je zobrazen v grafu 4.3. Podrobné srovnání tokenizátorů bude uvedeno v kapitole 5. Další grafy pro všechny použité tokenizátory a podrobné statistiky tohoto datasetu lze nalézt v souboru „Laptop-English.ipynb“.



**Obrázek 4.3** Porovnání dvou tokenizátorů na datasetu Recenze notebooků v angličtině, vlastní práce

#### 4.1.2.2 Recenze restaurací

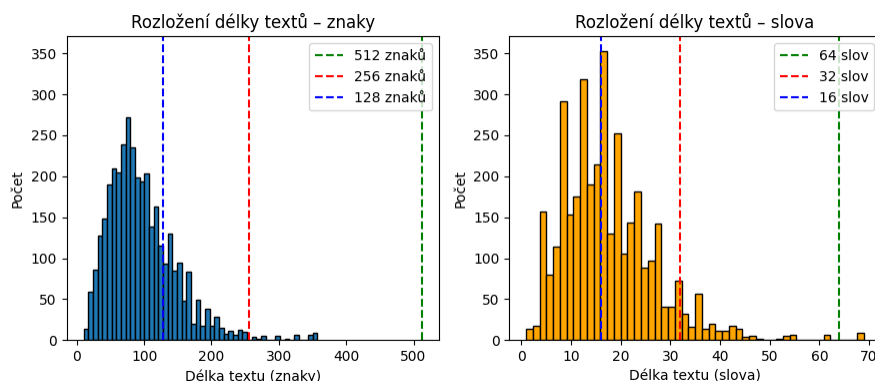
Data pro dataset recenzí restaurací byla také získána z portálu Hugging Face, kde je opět připravil Tom Aarsen, a jsou dostupná z [105]. Díky tomu, že datasety byly připraveny jedním autorem, je mezi daty zachována konzistence.

Tento dataset obsahuje 3602 vzorků, což ho činí podstatně větším než dataset recenzí notebooků. Vzorky jsou rozděleny na 2521 trénovacích, 720 validačních a 361 testovacích. Tento dataset obsahuje 1976 unikátních textů a 1268 unikátních aspektů. Rozdělení sentimentu v tomto datasetu je zobrazeno v grafu 4.1. Je zde patrné, že pozitivní data převyšují negativní téměř trojnásobně, což znamená, že modely trénované na těchto datech budou mít tendenci predikovat pozitivní sentiment. Proto se dá předpokládat, že celková přesnost bude vyšší než u modelů trénovaných na jiných datech. Průměrná délka textu v tomto datasetu je 96,5 znaku, což odpovídá přibližně 17,5 slovům. Nejdelší text má 357 znaků (69 slov). Celkové rozložení délky textu je znázorněno v grafu 4.4.

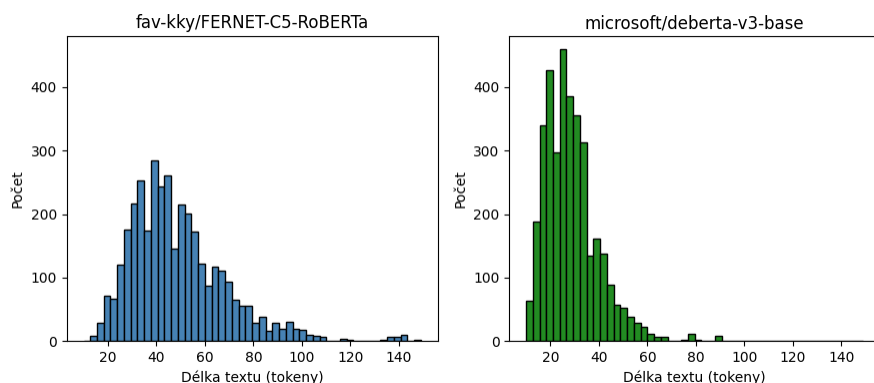
Mezi nejčastější aspekty v tomto datasetu patří: *food* (jídlo) a *service* (obsluha), které se vyskytují mnohem častěji než ostatní aspekty. Pro zjištění ideální délky vstupů pro model bylo provedeno testování různých tokenizátorů. Maximální délky textů po tokenizaci pro jednotlivé tokenizátory se pohybovaly mezi 90 a 149 tokeny. Zde je opět vcelku velký rozsah maximálních délek, zobrazený v grafu 4.5. Další grafy pro všechny použité tokenizátory a další statistiky týkající se tohoto datasetu jsou k dispozici ve zdrojovém kódu v souboru: „Restaurant-English.ipynb“.

#### 4.1.3 Český dataset recenzí restaurací

Tento dataset vznikl jako česká obdoba datasetu SemEval-2014. Autoři využili recenze restaurací z portálu [www.nejezto.cz](http://www.nejezto.cz), kde vybrali deset podniků s nejvyšším počtem hodnocení. Anotaci provedli tři rodilí mluvčí; podrobnosti



**Obrázek 4.4** Rozdělení délky textů v datasetu Recenze restaurací v angličtině, vlastní práce



**Obrázek 4.5** Porovnání dvou tokenizátorů na datasetu Recenze restaurací v angličtině, vlastní práce

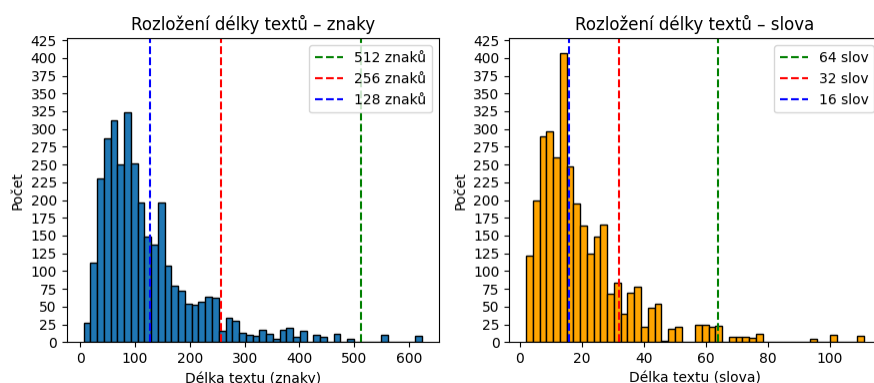
o postupu a pravidlech značení jsou popsány v práci [98]. Samotná data jsou dostupná v repozitáři [106].

Dataset obsahuje 3281 záznamů, rozdělených na 2296 trénovacích, 656 validačních a 329 testovacích. Z těchto záznamů je 1658 unikátních textů a 1150 unikátních aspektů. Rozdělení polarit sentimentu je zachyceno v grafu 4.1. Průměrná délka textu činí 124 znaků (přibližně 20,4 slova), nejdelší recenze má 623 znaků (111 slov). Distribuci délek recenzí ukazuje graf 4.6.

Nejčastěji se v anotacích vyskytují aspekty *jídlo*, *obsluha* a *restaurace*. Mezi deset nejfrekventovanějších patří také tvary *restauraci*, *jídla* či *Jídlo*. Různé pádové tvary ilustrují jednu z komplikací českého jazyka oproti angličtině. Do budoucna by bylo vhodné dataset normalizovat, případně při trénování využít lemmatizaci nebo stemming popsané v sekci 1.1.1.1.

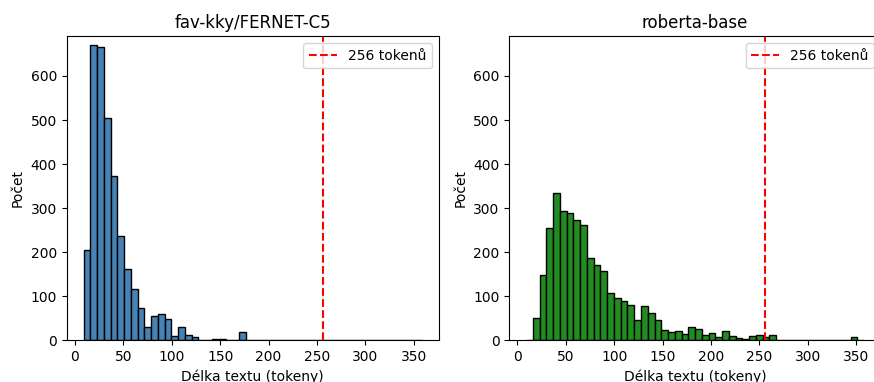
V tomto datasetu je patrný mnohem větší rozdíl mezi tokenizátory, než v předchozích případech. Maximální délky tokenizovaných sekvencí se pohybovaly v rozmezí od 176 do 358 tokenů. Tento výrazný rozdíl je způsoben





**Obrázek 4.6** Rozdělení délky textů v datasetu Recenze restaurací v češtině, vlastní práce

tím, že čeština má více znaků, se kterými mají anglické tokenizátory problém. Anglické tokenizátory jsou trénovány primárně na anglických textech a proto mají problém s češtinou, která obsahuje více diakritických znamének a složitější morfologii (např. více pádů). V grafu 4.7 je zobrazen největší rozdíl mezi dvěma tokenizátory, který ukazuje tento problém. Grafy pro všechny použité tokenizátory a další statistiky tohoto datasetu jsou k dispozici ve zdrojovém kódu v souboru „Restaurant-Czech.ipynb“.



**Obrázek 4.7** Porovnání dvou tokenizátorů na datasetu Recenze restaurací v češtině, vlastní práce

#### 4.1.4 Data z mediální domény

Původně měla být použita reálná data z archivu Newton Media. Analýza však ukázala, že dostupné korpusy jsou pro trénování jazykových modelů nevhodné – anotace aspektů jsou nevyvážené a počet tříd výrazně kolísá. Proto byl vytvořen *syntetický* dataset, který byl následně použit pro trénování modelů.

#### 4.1.4.1 Popis generování dat

Data byla vygenerována pomocí modelu Gemini 2.0 Flash [107]. Proces generování probíhal tak, že z reálných článků byly nejprve vybrány tři nejdůležitější aspekty. Každému z těchto aspektů byl následně přiřazen náhodný sentiment. Poté bylo vytvořeno shrnutí daného článku, přičemž byl kladen důraz na zachování přiřazeného sentimentu k daným aspektům. K extrakci aspektů byl použit prompt uvedený v 4.1. Pro generování shrnutí s konkrétním sentimentem k těmto aspektům byl následně použit druhý prompt, uvedený v 4.2. Přesné použití těchto promptů a celý proces generování lze nalézt v příloze v souboru „script.py“.

```
f"""Najdi {n} nejdůležitějších, vzájemně odlišných aspektů,
které se týkají hlavního tématu následujícího textu. Aspekt je
jednoduché podstatné jméno nebo fráze, která popisuje konkrétní
vlastnost nebo aspekt daného tématu. Například: „kvalita
produktu“, „cena“, „zákaznický servis“.
Napiš seznam aspektů, které jsou relevantní pro daný text.
Nepoužívej žádné úvodní fráze ani vysvětlení, pouze seznam
aspektů."""
```

■ **Výpis kódu 4.1** Ukázka promptu pro extrakci nejdůležitějších aspektů, vlastní práce

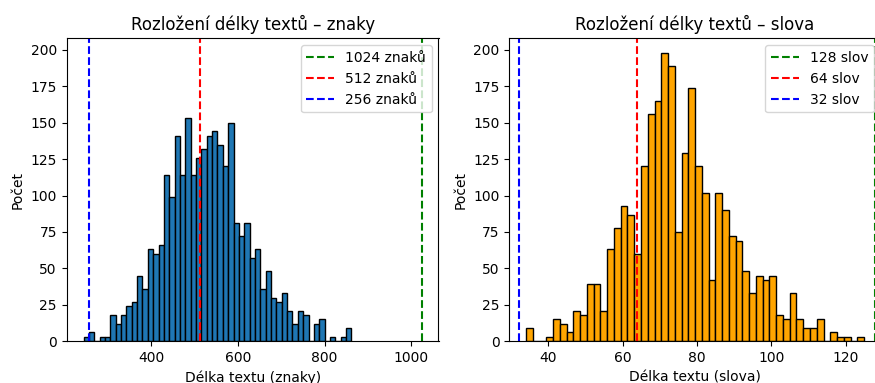
```
f"""Následující text se týká několika aspektů: {aspects_repr}.
Napiš souvislé shrnutí (max 200 slov), kde každý aspekt je
zmíněn a popsán tónem, který mu byl přiřazen (pozitivní/
negativní/neutrální). Snaž se o obecné shrnutí, kde jsou
aspekty mimochodem zmíněny.
Nezapomeň na žádný aspekt a dbej, aby výsledný odstavec působil
přirozeně. Pamatuji, že to má být shrnutí, ne analýza! I kdyby
byl aspekt v textu zmíněn například s jiným tónem než zadáno,
shrnutí by mělo být v souladu s tónem zadaným v úkolu a nikoli
s tónem v textu."""
```

■ **Výpis kódu 4.2** Ukázka promptu pro generaci shrnutí, kde se dbá na daný sentiment k daným aspektům, vlastní práce

#### 4.1.4.2 Popis samotného datasetu

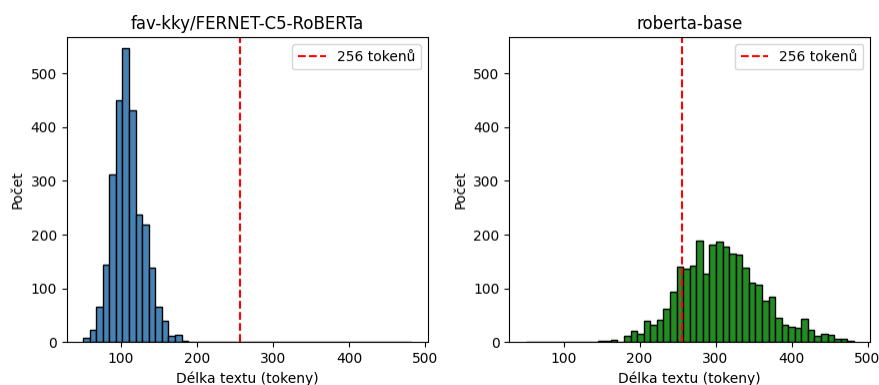
Po finálních úpravách obsahuje dataset 2709 záznamů, rozdělených na 1896 trénovacích, 542 validačních a 271 testovacích instancí. Celkem zahrnuje 903 unikátních textů a 1681 aspektů. Polaritní rozložení je vyvážené (graf 4.1),

což eliminuje zkreslení při trénování modelů. Průměrná délka textu činí 529,2 znaku, přibližně 75,5 slova, nejdelší text dosahuje 862 znaků (125 slov). Distribuci délek ukazuje graf 4.8. Oproti předchozím datasetům jsou texty podstatně delší, což odpovídá charakteru mediálních článků.



**Obrázek 4.8** Rozdělení délky textů v datasetu Mediální texty v češtině, vlastní práce

Nejfrekventovanějšími aspekty v celém korpusu jsou *coming out*, *dlouholetý vztah* a *příčina výpadku*; žádný z nich však nevyčnívá svou absolutní četností natolik, aby převážil nad ostatními. V tomto datasetu je rozdíl mezi tokenizátory ještě výraznější než u dříve popsaných sad – maximální délky tokenizovaných sekvencí se pohybovaly od 170 do 482 tokenů. Graf 4.9 ukazuje kontrast mezi dvojicí tokenizátorů stejné architektury RoBERTa [23]: česká varianta dosahuje svých nejdelších sekvencí přibližně na úrovni nejkratších sekvencí anglické verze, což dokládá, že anglické tokenizátory si s češtinou poradí podstatně hůře. Další grafy délek po tokenizaci a doplňkové statistiky k tomuto korpusu jsou k dispozici v příloženém notebooku „Media-Czech.ipynb“.



**Obrázek 4.9** Porovnání dvou tokenizátorů na datasetu Mediální texty v češtině, vlastní práce

## 4.2 Vybrané metody

Využití BERT modelů pro ABSA se zabývali autoři článku [54], kteří představili čtyři nová řešení zaměřená specificky na TASD a E2E-ABSA. Využívají vlastnost modelu BERT, konkrétně jeho struktura vstupu, díky čemuž je možné převést úkoly ABSA (ať už jakékoliv podúlohy) na klasifikaci dvojice vět (*sentence-pair classification*). Tento přístup bude testován i v této práci, kde bude porovnávána klasifikace dvojice vět s klasifikací jedné věty (*single-sentence classification*).

Při převodu ABSA úloh na klasifikaci dvojice vět se druhá věta používá jako doplňující informace k získání sentimentu z původního textu. Jedna věta je původní text a druhá slouží jako doplňující (*auxiliary sentence*). Tato pomocná věta může být formulována buď jako otázka (*QA*), nebo jako pseudo-věta (*NLI*). Tyto dva přístupy se pak dále dělí podle toho, zda bude odpověď polarita sentimentu (*M*), nebo odpověď typu ano/ne (*B*) [54].

### ■ QA-M

Doplňující věta je ve formě otázky, na kterou bude odpověď už daná polarita sentimentu. Příklad otázky: „Jaký je sentiment tohoto aspektu: {aspect}?", kde za {aspect} může být dosazen libovolný požadovaný aspekt z daného textu.

### ■ NLI-M

Doplňující věta v tomto případě nemá přísná pravidla a jedná se o tzv. pseudo-větu. Může jít například o větu jako: „aspekt: {aspect}“. Model pak vrací polaritu sentimentu na základě této věty.

### ■ QA-B

Doplňující věta je opět ve formě otázky, ale úloha je převedena na binární odpověď (ano/ne). Polarita sentimentu může mít tři možné hodnoty: „pozitivní“, „negativní“ a „neutrální“. V této metodě je třeba vygenerovat tři různé otázky pro jeden text, což znamená, že predikce musí být provedena třikrát, aby byly získány odpovědi pro všechny tři polarities. Příklad možných otázek: „Je sentiment pro aspekt: {aspect} pozitivní?“, „Je sentiment pro aspekt: {aspect} negativní?“ a „Je sentiment pro aspekt: {aspect} neutrální?“.

### ■ NLI-B

Opět jde o pseudo-větu, která však již obsahuje informaci o sentimentu. Opět bude potřeba vygenerovat tři věty pro jeden text. Příklad takových vět: „aspekt: {aspect}, sentiment: pozitivní“, „aspekt: {aspect}, sentiment: negativní“, „aspekt: {aspect}, sentiment: neutrální“. [54]

Jak již bylo zmíněno, v této práci budou použity dvě metody: jedna pro klasifikaci jedné věty a druhá pro klasifikaci dvojice vět. Pro klasifikaci dvojice vět bude z metod popsanych výše vybrána metoda *QA-M*. Tato metoda byla

zvolena na základě pozorování v práci [54], která ukázala, že metody typu *QA* dosahovaly lepších výsledků při predikci sentimentu. Verze *M* byla vybrána kvůli její jednoduchosti a menší náročnosti na trénování (u verze *B* by trénování trvalo třikrát déle). Popis implementace těchto metod bude uveden v sekci 4.4.3.1.

### 4.3 Vybrané modely

Tato část představuje všechny modely testované v rámci této práce a vysvětluje jejich rozdělení do dvou logických skupin. Mezi nimi bude vybírán nejvhodnější kandidát pro český mediální prostor a na každý z nich se aplikují obě metody popsané v sekci 4.2. Všechny modely jsou založeny na transformátorové architektuře a vycházejí z modelu BERT, který je rovněž součástí této práce.

Modely jsou rozděleny na *inovativní* a *jazykově adaptované*. První skupinu tvoří architektury, které přinášejí nové předtrénovací postupy či mechanismy (např. RoBERTa, DeBERTa-V3, ModernBERT). Do druhé skupiny patří verze již existujících modelů přetrénované pro konkrétní jazyková prostředí, zejména pro češtinu a další slovanské jazyky (SlavicBERT, Czert, RobeCzech, FER-NET).

Pro popis těchto modelů jsou používány následující parametry: *počet vrstev*  $L$  (počet transformátorových bloků), *velikost skrytého vektoru*  $H$  (velikost vektoru, který reprezentuje jednotlivé tokeny v každé vrstvě transformátoru) a *počet self-attention hlav*  $A$  (počet nezávislých mechanismů pozornosti, které model používá k výpočtu pozornosti pro každý token vzhledem k ostatním tokenům) [9]. Vypsání parametrů všech modelů lze nalézt v tabulce 4.1.

#### 4.3.1 Inovativní modely

Následující podsekcce popisují modely, které posouvají možnosti transformátorových architektur prostřednictvím nových trénovacích technik, modifikované pozornosti či efektivnější implementace. Tyto inovace mají za cíl zvýšit přesnost, zrychlit rozhodování a snížit paměťové nároky, což je klíčové zejména při zpracování rozsáhlých textových korpusů.

##### 4.3.1.1 BERT

Model BERT (zkratka pro *Bidirectional Encoder Representations from Transformers*) byl představen v práci [9] jako řešení pro široké spektrum úloh na úrovni věty a tokenu (sentence-level and token-level tasks). Předchozí modely byly zaměřeny pouze na jeden směr čtení textu, zatímco BERT, díky své *dvojcestné* (bidirectional) architektuře, dokáže zohlednit kontext z obou stran textu, což zlepšuje predikci tokenů. BERT je předtrénovaný model, který se

model	L	H	A	počet parametrů
bert-base-uncased	12	768	12	110 milionů
bert-large-uncased	24	1024	16	340 milionů
bert-base-multilingual-cased	12	768	12	110 milionů
roberta-base	12	768	12	125 milionů
xlm-roberta-base	12	768	12	270 milionů
xlm-roberta-large	24	1024	16	550 milionů
deberta-v3-base	12	768	12	184 milionů
ModernBERT-base	22	768	12	149 milionů
ModernBERT-large	28	1024	16	395 milionů
bert-base-bg-cs-pl-ru-cased	12	768	12	180 milionů
Czert-B-base-cased	12	768	12	110 milionů
robeczech-base	12	768	12	126 milionů
FERNET-C5	12	768	12	164 milionů
FERNET-C5-RoBERTa	12	768	12	124 milionů

■ **Tabulka 4.1** Parametry použitých modelů

nejprve trénuje na obecných datech a poté je možné jej doladit (fine-tune) na specifických úlohách s menšími datovými sadami.

BERT je postaven na transformátorové architektuře, která byla představena v sekci 1.2.2.1, využívající model encoder, popsán v sekci 1.2.3. Vstupy do modelu BERT mohou být jak jednotlivé věty, tak dvojice vět (například otázka a odpověď). Tato flexibilita rozšiřuje možnosti využití modelu, jak je tomu i v této práci. Pro předtrénování modelu BERT byly použity dva rozsáhlé datasety: BooksCorpus (800 milionů slov) a anglická Wikipedie (2,5 miliardy slov). Předtrénovací techniky zahrnovaly *Missing token prediction* a *Next sentence prediction*, které jsou podrobněji popsány v sekci 1.2.2.2. Podrobnější popis modelu BERT lze nalézt v díle [9].

V této práci byly použity dvě verze modelu BERT: *bert-base-uncased* [108] a *bert-large-uncased* [109]. Jedná se o stejné modely, lišící se pouze velikostí. Větší modely jsou schopny zachytit složitější vlastnosti textu, ale jsou výpočetně náročnější. [9] Zvoleny byly *uncased* verze, protože patří k nejpoužívanějším anglickým modelům BERT a v praxi se nasazují častěji než jejich *cased* protějšky (*bert-base-cased* [110] a *bert-large-cased* [111]).

#### 4.3.1.2 mBERT

Model *multilingual BERT* (mBERT) vychází ze stejné architektury jako původní anglický BERT; liší se pouze trénovacími daty. Byl trénován na Wikipediích ve 100 jazycích s největším počtem článků. Rozložení dat mezi jazyky je však nerovnoměrné – například angličtina tvoří přibližně 21 % veškerých trénovacích dat. Aby i jazyky s omezenými zdroji (*low-resource* jazyky) získaly

dostatečné zastoupení, autoři použili vyvážené vzorkové (sample) schéma. Podrobné informace jsou uvedeny v repozitáři projektu [112].

mBERT nemá *large* variantu, proto je v této práci použit model *bert-base-multilingual-cased* [113]. Verze *cased* je doporučena tvůrci modelu, protože opravuje normalizační problémy u mnoha jazyků, speciálně je doporučena pro jazyky, které nepoužívají latinku (většinou je lepší i pro jazyky, které latinku používají). [112]

#### 4.3.1.3 RoBERTa

Model RoBERTa (*Robustly Optimized BERT Pretraining Approach*) vznikl na základě zjištění, že původní BERT je ve fázi předtrénování nedostatečně využit. Autoři proto navrhli několik úprav předtrénovacího postupu:

1. delší trénování na rozsáhlejších datech,
2. odstranění *Next Sentence Prediction*,
3. trénování na delších sekvencích,
4. dynamickou změnu maskovacího vzoru v průběhu trénování.

Díky uvedeným úpravám dosahuje RoBERTa ve srovnání s původním modelem BERT-em výrazně lepších výsledků napříč různými jazykovými úlohami. Podrobné zdůvodnění jednotlivých vylepšení a jejich vliv na výkon modelu je rozebráno v původním článku [23].

V této práci je použit model *roberta-base* [114]. RoBERTa se nečlení na varianty *cased* a *uncased*, takže nebylo nutné volit mezi různými verzemi.

#### 4.3.1.4 XLM-RoBERTa

XLM-RoBERTa je vícejazyčný model navržený jako nástupce mBERT pro vícejazyčné úlohy. Vznikl kombinací optimalizovaného předtrénovacího postupu z modelu RoBERTa a techniky předtrénování XLM (*cross-lingual language models*, vícejazyčné jazykové modelování) představené v práci [115]. Výsledkem je model schopný pracovat se 100 jazyky, který překonává předešlé nejlepší modely. Podrobnosti o architektuře a trénování jsou popsány v [116].

V této práci jsou použity obě dostupné velikosti: *xlm-roberta-base* [117] a *xlm-roberta-large* [118]. Jde o jediný vícejazyčný model v této práci, který je k dispozici i ve variantě *large*.

#### 4.3.1.5 DeBERTa-V3

Modelová architektura DeBERTa (*Decoding-enhanced BERT with disentangled attention*) rozšiřuje možnosti BERT a RoBERTa dvěma hlavními inovacemi. První z nich je rozdělená pozornost (*disentangled attention*): každé slovo je

reprezentováno dvojicí vektorů – jeden nese obsahovou informaci, druhý relativní pozici – a váhy pozornosti se počítají odděleně pro obsahové a polohové složky. Druhou novinkou je vylepšený maskovací dekodér, jenž do předtrénování začleňuje i absolutní pozice tokenů. Podrobnosti těchto inovací uvádí práce [119].

Tato práce využívá novější variantu *DeBERTa-V3*, jež nahrazuje klasickou techniku předtrénování predikce chybějícího tokenu (*missing token prediction*) technikou predikce vyměněného tokenu (*replaced token detection*). Také je zde představena nová metoda *gradient-disentangled embedding sharing*, která zlepšuje efektivnost trénování a celkovou kvalitu modelu. Při stejném předtrénovacím režimu jako původní DeBERTa tato vylepšení ukazují zlepšení na širokém spektru NLU úloh. Podrobnější popis těchto vylepšení lze nalézt v díle [120].

Použitým modelem je *deberta-v3-base* [121]. Existuje i vícejazyčná varianta mDeBERTa-V3, ta však pokrývá pouze 16 jazyků a čeština mezi nimi chybí, proto není v této práci použita.

#### 4.3.1.6 ModernBERT

ModernBERT představuje nejnovější encoder-only model vycházející z původního BERT-u. Autoři do klasické transformátorové architektury postupně začlenili řadu ověřených novinek a přidali vlastní úpravy zaměřené na efektivitu jak z hlediska rychlosti, tak paměťové náročnosti. Současně optimalizovali implementaci pro trénování na GPU. Podrobný popis všech změn je uveden v práci [24].

Ve srovnání s dřívějšími modely (např. DeBERTa) je ModernBERT výrazně úspornější a typicky alespoň dvakrát rychlejší; v některých úlohách autoři uvádějí až čtyřnásobné zrychlení při současném snížení paměťových nároků. Další zásadní předností je maximální kontextová délka 8192 tokenů, tedy téměř šestnáctinásobek běžného limitu 512 tokenů. Přehled měřených vlastností a praktická srovnání shrnuje blogový článek [22].

V této práci jsou otestovány obě velikosti modelu, *ModernBERT-base* [122] a *ModernBERT-large* [123]. Vícejazyčná varianta zatím neexistuje; kdyby byla k dispozici, nabízela by ideální kombinaci dlouhého kontextu a mnohojazyčné podpory pro analýzu rozsáhlých mediálních textů.

#### 4.3.2 Jazykově adaptované modely

Tato podsekcce se věnuje variantám, které nepřinášejí nové architektonické prvky, ale rozšiřují stávající modely na méně zastoupené jazyky – zde především na češtinu a další slovanské jazyky. Jejich hlavním přínosem je tedy jazyková adaptace, využívají osvědčené inovace z předchozí skupiny a převádějí je do jazykového prostředí, kde původní anglické a vícejazykové modely nedosahují optimální výkonnosti.



#### 4.3.2.1 SlavicBERT

SlavicBERT patří mezi první specializované varianty BERT-u zaměřené na slovanské jazyky (češtinu, ruštinu, polštinu a bulharštinu). Autoři vyšli z modelu *mBERT* a dále jej doladili na korpusech výhradně v těchto čtyřech jazycích, protože trénovat novou architekturu od začátku by bylo výpočetně i finančně nákladné. Původně byl model navržen pro rozpoznávání vícejazyčných pojmenovaných entit (NER, named entity recognition), nicméně je použitelný i pro další úlohy zpracování přirozeného jazyka [124].

V této práci je využita verze *bert-base-bg-cs-pl-ru-cased* [125], která je na portálu Hugging Face dostupná jako jediná oficiálně zveřejněná varianta.

#### 4.3.2.2 Český BERT – Czert

Cílem autorů modelu Czert bylo vytvořit model vycvičený výhradně na českých datech. Na rozdíl od SlavicBERT-u zvolili trénování *od nuly*, aby se model učil jazyk čistě z českého korpusu. K předtrénování použili přibližně 36 GB textu. Architektura i postup kopírují původní BERT, jedinou úpravu vyžadovala technika *Next Sentence Prediction*: vzhledem ke krátkým textovým blokům často chyběla vhodná následující věta, což autoři řešili v detailech popsanych v práci [100].

Vznikly dvě varianty: *Czert-B* založená na BERT-u a lehčí *Czert-A* vycházející z ALBERT-u. Pro tuto práci je zvolen model *Czert-B-base-cased* [126], jehož rozměry odpovídají mBERT-u a usnadňují tak férové porovnání s ostatními modely.

#### 4.3.2.3 Český RoBERTa – RobeCzech

Na dosavadní český model založený na architektuře BERT (Czert) navázal model RobeCzech, jenž vychází z architektury RoBERTa. Předtrénování proběhlo na souboru českých korpusů o celkovém rozsahu převyšujícím čtyři miliardy tokenů a k tréninku byl použit oficiální kódový rámec pro RoBERTa. Následně byl RobeCzech doladěn a vyhodnocen na pěti úlohách z oblasti NLP, mezi nimiž byla i klasická (ne-aspektová) analýza sentimentu. Ve všech sledovaných úlohách překonal ostatní modely stejné velikostní kategorie. Podrobný popis korpusu, parametrů trénování i výsledků benchmarku je uveden v práci [102].

V této práci se používá *robeczech-base* [127]; jedná se o jedinou dosud publikovanou variantu tohoto modelu, další verze zatím nebyly vytvořeny.

#### 4.3.2.4 Český BERT a RoBERTa – FERNET

Název FERNET zastřešuje dvojici českých modelů, jejichž cílem bylo rozšířit portfolio dostupných jazykových modelů pro češtinu. Autoři zvolili jiné předtrénovací korpusy než u předchozích modelů Czert a RobeCzech – konkrétně vlastní korpusy *Czech News* a *C5*, jenž představuje českou obdobu anglického

C4 [128]. Na těchto datech vznikly dvě varianty: *FERNET-C5*, vycházející z architektury BERT, a *FERNET-News*, založená na architektuře RoBERTa. Oba modely byly po jemném doladění srovnány na dvou úlohách z oblasti NLP, z nichž jedna byla opět klasická analýza sentimentu; v tomto srovnání dosáhl FERNET-C5 nejlepších výsledků mezi dostupnými českými modely. Detailní popis přípravy dat, parametrů trénování a dosažených skóre přináší studie [103].

V této práci budou využity dvě varianty: *FERNET-C5* [129] a *FERNET-C5-RoBERTa* [130]. Druhý model, ačkoli nebyl součástí původního článku, je model založený na architektuře RoBERTa, který byl předtrénován jak na datasetu C5, tak na datasetu českých novinek.

## 4.4 Popis přístupu

V předchozích částech byly podrobně představeny použité datasety, metody i jazykové modely. Tato sekce vysvětluje, *jak* byly všechny tyto komponenty propojeny v praxi. Popisuje trénovací prostředí, klíčové knihovny a důležité fragmenty kódu, které zajišťují trénování a vyhodnocení modelů.

### 4.4.1 Trénovací prostředí

Veškeré trénování probíhalo v prostředí Google Colab [131] – hostovaném Jupyter notebooku – který poskytuje zdarma přístup k výpočetním zdrojům, včetně GPU a TPU. Tato varianta je pro akademické účely výhodná, protože odstraňuje nutnost vlastnit výkonný hardware.

V Colabu (ve verzi použité během trénování) byl využit virtuální stroj s 12,7 GB systémové RAM a grafickou kartou *NVIDIA Tesla T4* [132] s 15 GB paměti. Přestože T4 není nejvýkonnější čip v nabídce Colabu, poskytla dostatečný výkon. Všechny modely bylo možné natrénovat bez placeného tarifu a bez nutnosti zkracovat dávky či redukovat maximální délku vstupu.

### 4.4.2 Použité knihovny

Tato podsekce stručně charakterizuje knihovny využité při implementaci trénování. Všechny moduly jsou součástí ekosystému *Hugging Face* [10] nebo patří mezi standardní nástroje pro práci s daty a hluboké učení.

- **Datasets** [133]

Knihovna *Datasets* usnadňuje stahování, ukládání a transformaci datových sad. V rámci práce je používána datová struktura *Dataset*, do níž jsou načteny trénovací, validační a testovací podmnožiny.

- **pandas** [134] a **NumPy** [135]

Základní nástroje pro manipulaci s tabulkovými daty (*pandas*) a efektivní matematické operace (*NumPy*).

- **Transformers** [136]

Klíčová knihovna poskytující rozhraní k předtrénovaným modelům, tokenizátorům a trénovacím nástrojům. Umožňuje rychlé jemné doladění (*fine-tuning*) modelů i jejich trénování od začátku.

- **Evaluate** [137]

Modul s implementacemi vyhodnocovacích metrik; ve spojení s *Transformers* zajišťuje jednotný rámec pro sledování výkonu při trénování.

- **PyTorch** [138]

Framework hlubokého učení použitý jako backend pro trénování a učení neuronových sítí.

### 4.4.3 Popis částí kódu

Tato podsekce se zaměřuje na klíčové fragmenty zdrojového kódu, který byl použit při trénování jednotlivých modelů. Veškeré trénování probíhalo pomocí stejného trénovacího kódu, což zajišťuje konzistenci a reprodukovatelnost výsledků. Dále jsou vysvětleny parametry použité při trénování a upozorněno na odlišnosti mezi oběma aplikovanými metodami. Podsekce rovněž ukazuje funkci, která pro zadaný text a aspekt vrátí odpovídající sentiment. Kompletní skripty pro všechny modely jsou k dispozici v příloze této práce.

#### 4.4.3.1 Tokenizace

Knihovna *Transformers* umožňuje načíst všechny podporované tokenizátory jedním příkazem a okamžitě je použít. Funkce uvedené v kódech 4.3 a 4.4 zajišťují tokenizaci pro obě zvolené metody; liší se pouze způsobem předání vstupu, ostatní argumenty zůstávají totožné.

- **Základní metoda** (ukázka 4.3): tokenizátor obdrží jediný řetězec složený z *aspect* a *text*.
- **Metoda QA-M** (ukázka 4.4): první věta obsahuje původní text, druhá věta je dotaz typu „Jaký je sentiment tohoto aspektu: {aspect}?“ V anglické variantě je otázka přeložena do angličtiny.

Jak ukazuje analýza dat v sekci 4.1, většina textů se po tokenizaci vejde do 256 tokenů. Výjimkou jsou ale české datasety zpracované tokenizátory netrénovanými na češtině, kde délka tokenizovaných textů byla delší než 256 tokenů. Jenže takové modely nejsou relevantní, neboť nejsou pro češtinu optimalizovány a nejsou od nich očekávané kvalitní výsledky na českých datech. Z tohoto důvodu a k zachování konzistence při trénování různých modelů je maximální délka tokenizovaného textu omezena na *256 tokenů*.

```
def tokenize_function(example):
    combined_input = f"aspect: {example['aspect']} text: \
{example['text']}"

    encoding = tokenizer(
        combined_input,
        truncation=True,
        padding="max_length",
        max_length=256
    )
    return encoding
```

■ **Výpis kódu 4.3** Ukázka funkce tokenizace u základní metody, vlastní práce

```
def tokenize_function(example):
    question_string = f"What is the sentiment of aspect: \
{example['aspect']}?"

    encoding = tokenizer(
        text=example["text"],
        text_pair=question_string,
        truncation=True,
        padding="max_length",
        max_length=256
    )
    return encoding
```

■ **Výpis kódu 4.4** Ukázka funkce tokenizace u metody QA-M, vlastní práce

#### 4.4.3.2 Vyhodnocovací metrika

Jako vyhodnocovací metrika během trénování a pro výběr nejlepší epochy bude použita *přesnost* (accuracy). S využitím knihovny *Evaluate* bude přesnost vypočítána pomocí funkcí uvedenou v ukázce kódu 4.5, která je následně předána objektu *Trainer* použitému pro samotné trénování.

#### 4.4.3.3 Trénovací struktura

Proces trénování je řízen řadou parametrů, které ovlivňují jeho průběh. Aby byla mezi modely zachována srovnatelnost, jsou v celé práci používány totožné hodnoty argumentů; výkonnější hardware by umožnil ladění více kombinací a podrobnější analýzu jejich vlivu na výsledky.

Níže jsou popsány klíčové hyperparametry objektu *TrainingArguments*. Úplný seznam všech podporovaných parametrů je uveden v oficiální doku-

```
metric = evaluate.load("accuracy")
def compute_metrics(eval_pred):
    logits, labels = eval_pred
    predictions = np.argmax(logits, axis=-1)
    return metric.compute(predictions=predictions,
                           references=labels)
```

■ **Výpis kódu 4.5** Ukázka funkce pro výpočet přesnosti, vlastní práce

mentaci [139]. Ukázka kódu 4.6 prezentuje konkrétní konfiguraci použitou při trénování všech modelů použitých v této práci.

- **learning\_rate** – učicí parametr, předávaný do optimalizátoru AdamW;
- **num\_train\_epochs** – počet trénovacích epoch;
- **weight\_decay** – regularizační složka omezující přeučení (over-fitting);
- **save\_total\_limit** – maximální počet uložených checkpointů;
- **warmup\_ratio** – podíl úvodních kroků, v nichž se *learning rate* lineárně zvyšuje z 0 na cílovou hodnotu.

```
training_args = TrainingArguments(
    output_dir="./results",
    learning_rate=3e-5,
    per_device_train_batch_size=16,
    per_device_eval_batch_size=16,
    num_train_epochs=10,
    weight_decay=0.01,
    eval_strategy="epoch",
    save_strategy="epoch",
    save_total_limit=3,
    logging_dir="./logs",
    logging_steps=200,
    warmup_ratio=0.06,
    load_best_model_at_end=True,
    metric_for_best_model="accuracy"
)
```

■ **Výpis kódu 4.6** Ukázka struktury TrainingArguments, vlastní práce

Tyto argumenty jsou předány do objektu *Trainer*, ukázaný v 4.7, jenž zajišťuje vlastní trénování modelu. V parametru *compute\_metrics* je předána funkce pro výpočet zvolené metriky (přesnosti). Samotné trénování se spustí voláním `trainer.train()`.

```

trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=train_dataset,
    eval_dataset=val_dataset,
    compute_metrics=compute_metrics
)

```

■ **Výpis kódu 4.7** Ukázka struktury Trainer, vlastní práce

#### 4.4.3.4 Predikace sentimentu

Jakmile je model natrénován a připraven k použití, je potřeba vytvořit funkci, která umožní jeho aplikaci na nový text. Pro predikci sentimentu pro jeden text a jeden aspekt bude použita funkce uvedená v 4.8. V této funkci se nejprve provede tokenizace vstupu pomocí stejného tokenizátoru, jaký byl použit při trénování modelu. Poté je tokenizovaný vstup předán do modelu, který vrátí vektor s výsledky pro každou třídu, přičemž pro predikci sentimentu bude vybrána třída s nejvyšší hodnotou.

Ukázka funkce pro modely s klasickou metodou je uvedena v 4.8. Pro metodu *QA-M* je však nutné provést tokenizaci podle postupu, ukázaného v 4.4.

```

def predict_sentiment(model, tokenizer, aspect, text):
    combined_input = f"aspekt: {aspect} text: {text}"

    encoding = tokenizer(
        text=combined_input,
        return_tensors="pt",
        truncation=True,
        padding="max_length",
        max_length=256
    )
    with torch.no_grad():
        outputs = model(**encoding)

    logits = outputs.logits
    print(logits)
    predicted_class_id = logits.argmax(dim=1).item()
    return id2label[predicted_class_id]

```

■ **Výpis kódu 4.8** Ukázka funkce na predikci sentimentu, vlastní práce

## 4.5 Vyhodnocovací metriky a praktické testy

Předchozí sekce byly zaměřeny na trénování jednotlivých modelů – od popisu použitých metod, přes výběr modelů a datových sad, až po samotný postup trénování. Jakmile však všechny modely budou natrénovány, je nutné vybrat ten nejvhodnější, který bude schopen nejlépe řešit úlohy ABSA.

K tomuto účelu se v této práci uplatňují dva navzájem se doplňující přístupy. Prvním jsou klasické vyhodnocovací metriky: vedle již zmíněné přesnosti se sleduje také makro-F1 a vážené F1 skóre (weighted F1). Druhým přístupem jsou takzvané *praktické testy*, které prověřují schopnost modelů zobecnit na širší tematické okruhy a pracovat i s doménami, jež se v trénovacích datech nevyskytovaly.

### 4.5.1 Vyhodnocovací metriky

Vyhodnocovací metriky slouží jako obecný ukazatel kvality modelu, přičemž jsou počítány na testovací množině, tedy na datech, která nebyla použita během samotného trénování. Na základě těchto metrik lze vybrat nejlepší model ze všech natrénovaných modelů na stejném datasetu, protože bude použit stejný dataset na testování kvality.

V této práci budou použity tři metriky: přesnost, macro F1 a vážená F1. Každá z těchto metrik měří jiný aspekt výkonu modelu, a proto je vhodné vyhodnotit všechny tři pro komplexnější popis jednotlivých modelů.

- **Přesnost (accuracy)** – podíl správně klasifikovaných vzorků z celkového počtu zpracovaných případů [140];
- **F1 skóre** – harmonický průměr přesnosti (precision) a výtěžnosti (recall).
  - **Macro F1** – nevážený průměr F1 skóre u všech tříd (label). Tato metrika nezohledňuje nerovnováhu mezi třídami;
  - **Vážená F1** – vážený průměr F1 skóre u všech tříd (label), který bere v úvahu nerovnováhu mezi třídami [141].

### 4.5.2 Praktické testy

Klasické metriky, jako je přesnost či F1 skóre, shrnují výkon modelu jediným číslem a samy o sobě neodhalí, jak si model poradí s konkrétními situacemi. Proto jsou do této práce zařazeny čtyři sady krátkých textů pocházejících z různých domén. Každý text je připraven v angličtině i češtině, aby bylo možné otestovat modely v obou jazycích. V jedné větě se vždy nachází několik aspektů, což umožňuje ověřit, zda model správně rozezná více aspektů v téže větě a přiřadí každému aspektu odpovídající polaritu.

### 4.5.2.1 Recenze notebooků L

L-sada obsahuje ukázkou z domény, na níž byla část modelů přímo trénována, a slouží tedy jako kontrola, zda si tyto modely udrží výhodu i v ručně vytvořeném příkladu. Zároveň umožňuje porovnat jejich výstup s modely, které tréninková data z této oblasti neviděly, a posoudit tak vliv doménového přizpůsobení.

#### ■ Angličtina

„The **power** is good, but the **display** is blurry and dark. The **battery** was a real surprise, I don't need to charge it during the day!“

- **L1:** *power* – pozitivní,
- **L2:** *display* – negativní,
- **L3:** *battery* – pozitivní.

#### ■ Čeština

„**Výkon** je dobrý, ale **displej** je rozmazaný a tmavý. **Baterie** byla skutečným překvapením, nemusím ji přes den nabíjet!“

- **L1:** *výkon* – pozitivní,
- **L2:** *displej* – negativní,
- **L3:** *baterie* – pozitivní.

### 4.5.2.2 Recenze restaurací R

Stejně jako L-sada pokrývala jednu z trénovacích domén, R-sada se zaměřuje na recenze restaurací, druhou trénovací doménu. Umožní tak porovnat, jak se jednotlivé modely chovají v další relevantní oblasti a do jaké míry dokážou zobecnit znalosti mezi různými tématy.

#### ■ Angličtina

„When we walked into the restaurant, the **staff** were very rude and our **table** hadn't been cleaned. On the other hand, the **food** was really tasty and the **beer** was delicious.“

- **R1:** *staff* – negativní,
- **R2:** *table* – negativní,
- **R3:** *food* – pozitivní,
- **R4:** *beer* – pozitivní.



#### ■ Čeština

„Když jsme vešli do restaurace, **personál** byl velmi nepříjemný a náš **stůl** nebyl uklizený. Na druhou stranu bylo **jídlo** opravdu chutné a **pivo** bylo vynikající.“

- **R1:** *personál* – negativní,
- **R2:** *stůl* – negativní,
- **R3:** *jídlo* – pozitivní,
- **R4:** *pivo* – pozitivní.

#### 4.5.2.3 Mediální texty M

Protože hlavním cílem práce je doporučit model vhodný pro mediální doménu, M-sada ověřuje, zda si modely natrénované převážně na recenzích poradí i s krátkou zpravodajskou zprávou. V angličtině prověří zobecnění mimo tréninkové domény; v češtině umožní srovnat modely trénované na restauracích s těmi, které využily archiv Newton Media. Jména firem jsou smyšlená, aby nedošlo k poškození reálných subjektů.

#### ■ Angličtina

„At EXPO 2025 in **Osaka**, many companies introduced their latest products. **Saremi** successfully introduced a new laptop that visitors loved, whereas **Babtel**'s new phone was a disaster, it didn't even work!“

- **M1:** *Osaka* – neutrální,
- **M2:** *Saremi* – pozitivní,
- **M3:** *Babtel* – negativní.

#### ■ Čeština

„Na výstavě EXPO 2025 v **Ósace** představilo mnoho firem své nejnovější produkty. Společnost **Saremi** úspěšně uvedla nový notebook, který si návštěvníci oblíbili, zatímco nový telefon firmy **Babtel** byl katastrofa, vůbec nefungoval!“

- **M1:** *Ósace* – neutrální,
- **M2:** *Saremi* – pozitivní,
- **M3:** *Babtel* – negativní.

#### 4.5.2.4 Obecná doména O

Závěrečná sada prověřuje, jak si modely poradí s neutrálním, každodenním textem, který nepochází z žádné tréninkové domény. Slouží tedy jako test skutečné obecnosti.

**■ Angličtina**

„Today was a really good **day**, Anna asked me out for a date! Unfortunately the **weather** was awful, so we had to cancel it.“

- **O1:** *day* – pozitivní,
- **O2:** *weather* – negativní.

**■ Čeština**

„Dneska byl opravdu skvělý **den**, Anička mě pozvala na rande! Bohužel **počasí** bylo hrozné, takže jsme to museli zrušit.“

- **O1:** *den* – pozitivní,
- **O2:** *počasí* – negativní.

## Výsledky a porovnání

V této kapitole jsou prezentovány výsledky trénování všech modelů, přičemž během testování v jednotlivých doménách budou modely vzájemně porovnány podle jejich schopnosti správně predikovat sentiment a zmapovat danou doménu. Pro každou doménu bude vybrán nejlepší model, který vykázal nejlepší výsledky. Následně budou porovnány výsledky mezi jednotlivými doménami a jazyky, aby bylo otestováno, jaké modely měli konzistentní výsledky i v jiných doménách. Kromě toho budou porovnány obě použité metody trénování, aby se zjistilo, zda některá z nich vykazovala konstantně lepší výsledky než druhá.

Na závěr kapitoly bude uvedeno shrnutí výsledků trénování. Tento přehled bude zahrnovat výběr nejlepších modelů pro každý jazyk na základě dosažených výsledků. Také proběhne doporučení pro mediální texty v českém jazyce. Dále budou otestovány známé velké jazykové modely (*Large Language Models, LLM*), které v této práci nebyly použity, aby byly porovnány s menšími modely, použitými v této práci. Praktické testy provedené na těchto modelech zhodnotí jejich schopnost správně predikovat sentiment bez potřeby speciálního dotrénování (*fine-tuning*) pouze s použitím jednoho promptu. Nakonec bude nastíněn možný výhled do budoucna, jakým směrem by se mohl tento výzkum v oblasti aspektové analýzy sentimentu posunout.

## 5.1 Výsledky

V této sekci budou uvedeny výsledky pro jednotlivé domény. Nejprve budou prezentovány nejlepší modely, ve velikostech *base* a *large*, pro každou doménu a následně se shrnou obecné výsledky trénování napříč všemi modely. Dále budou porovnány výsledky praktických testů, aby se zjistilo, zda modely s nejlepšími výsledky na testovacích datech dosahovaly také nejlepších výsledků v praktických testech. U každé domény bude ukázána menší tabulka těch nejlepších modelů. Podrobnější tabulky obsahující všechny informace o trénování,

výsledcích na testovacích datech a praktických testech pro všechny modely jsou uvedeny v příloze A.

### 5.1.1 Recenze notebooků – angličtina

Nejlepší výsledky byly dosaženy nejnovějšími modely. Vítězem se stal model typu ModernBERT, a to jak verze *base*, tak verze *large*, které dosáhly nejlepších výsledků ve všech měřených metrikách na testovacích datech. U verze *base* měl lepší výsledky model s metodou *QA-M*, zatímco u verze *large* to byl model s klasickou metodou. Z výsledků v tabulce 5.1 je patrné, že mezi těmito dvěma velikostmi nebyl velký rozdíl, lišily se pouze o desetiny procenta. Model DeBERTa vykázal také velmi dobré výsledky, které byly dokonce lepší než ModernBERT s klasickou metodou.

model	přesnost	v. F1	m. F1
deberta-v3-base	<b>82,76</b>	<b>82,29</b>	<b>80,60</b>
deberta-v3-baseQA	81,90	81,67	<b>80,31</b>
ModernBERT-base	81,47	81,04	79,23
<b>ModernBERT-baseQA</b>	<b>84,48</b>	<b>84,19</b>	<b>82,65</b>
<b>ModernBERT-large</b>	<b>84,91</b>	<b>84,49</b>	<b>82,97</b>
ModernBERT-largeQA	<b>82,33</b>	<b>81,79</b>	79,86

■ **Tabulka 5.1** Tabulka nejlepších modelů pro dataset Recenze notebooků v angličtině ukazující vyhodnocovací metriky na testovacích datech

Z podrobnější tabulky A.1 je zřejmé, že modely vykazovaly malé rozpětí výsledků, testovací přesnost se pohybovala mezi 75,43 % a 84,91 %. Tento rozdíl je způsoben především vícejazyčnou verzí modelu BERT, mBERT, který vykázal horší výsledky o více než 4 %. Bez něho se výsledky pohybovaly mezi 79,31 % a 84,91 %. Tabulka A.2 ukazuje výsledky modelů trénovaných na češtině na anglických datech. Jak je patrné, tyto modely měly podstatně horší výsledky, přičemž testovací přesnost se pohybovala mezi 65,09 % a 72,25 %, což je téměř o 10 % horší než u modelů trénovaných na anglických datech.

Praktické testy odhalily dva modely, které predikovaly všechny hodnoty správně: *roberta-base* a *ModernBERT-large*, oba využívající metodu QA-M. I když tyto modely neměly nejlepší výsledky na testovacích datech (ani v rámci vážené F1, ani macro F1), dokázaly jako jediné správně predikovat sentiment ve všech praktických testech. Ostatní modely uvedené v tabulce 5.1 měly 11 správných odpovědí z 12. Důvodem, proč modely neměly dokonalé výsledky, byl test *M1*, který se zaměřuje na neutrální sentiment. Tento sentiment byl v datasetu zastoupen pouze 20 % (viz graf 4.1), což vedlo k podcenění jeho predikce během trénování (proto je macro F1 u většiny modelů nižší než přesnost). Výsledky všech modelů na praktických testech jsou v tabulkách A.3 a A.4.

### 5.1.2 Recenze restaurací – angličtina

Nejlepší výsledky opět dosáhly nejnovější modely. Pro modely *large* to byl opět ModernBERT, zatímco u modelů *base* dosáhl nejlepších výsledků model typu RoBERTa, který měl dokonce lepší výsledky než *ModernBERT-large*. Oba tyto modely měly nejlepší skóre ve všech měřených metrikách a oba byly natrénovány pomocí klasické metody. Tabulka 5.2 ukazuje nejlepší modely pro tento dataset. Modely DeBERTa měly také velmi dobré výsledky, ale na RoBERTa neuspěly. Model BERT ve verzi *large* dosáhl solidních výsledků, avšak v porovnání s novějšími modely zaostával.

model	přesnost	v. F1	m. F1
bert-large-uncased	84,76	84,80	80,13
bert-large-uncasedQA	83,93	83,77	79,25
<b>roberta-base</b>	<b>86,70</b>	<b>86,72</b>	<b>83,02</b>
roberta-baseQA	85,04	84,83	79,94
deberta-v3-base	<b>85,60</b>	<b>85,74</b>	<b>81,51</b>
deberta-v3-baseQA	84,21	84,51	79,57
<b>ModernBERT-large</b>	<b>86,15</b>	<b>85,74</b>	<b>81,09</b>
ModernBERT-largeQA	<b>85,87</b>	<b>85,74</b>	<b>81,08</b>

**Tabulka 5.2** Tabulka nejlepších modelů pro dataset Recenze restaurací v angličtině ukazující vyhodnocovací metriky na testovacích datech

Z podrobnější tabulky A.5 je patrné, že čistě anglické modely dosahovaly velmi podobných výsledků v rozmezí 80,33 % – 86,70 %. Vícejazyčné modely (mBERT a XLM-RoBERTa) vykazovaly horší výsledky než jejich čistě anglické varianty. Tabulka A.6 ukazuje výsledky trénování českých (slovanských) modelů, které opět měly horší výsledky než modely trénované na anglických datech. Testovací přesnost těchto modelů se pohybovala mezi 74,24 % a 77,84 %, což je o 6 % až 9 % horší než u modelů trénovaných na anglických datech. Zajímavé je, že rozdíl mezi přesností (nebo váženou F1) a macro F1 byl u jednotlivých modelů poměrně výrazný. Jelikož macro F1 testuje každou třídu (tedy každý sentiment) zvlášť, vykazuje horší výsledky u nevyvážených datasetů. Tento dataset obsahuje 60 % dat s pozitivním sentimentem (viz graf 4.1), což způsobuje, že metriky přesnosti a vážené F1, které se zaměřují na každý vzorek jednotlivě, vykazují vyšší hodnoty než macro F1.

Modely trénované na doméně recenzí restaurací dosáhly mnohem lepších výsledků v praktických testech než modely z předchozí domény. Až 11 modelů mělo všechny odpovědi správné, včetně těch uvedených v tabulce 5.2. Dokonce i modely trénované na češtině měly uspokojivé výsledky, přičemž dva modely dosáhly 11 správných odpovědí z 12. I přes nevyváženost tohoto datasetu dosáhly modely na praktických testech vysoké úspěšnosti. Výsledky všech modelů na praktických testech jsou zobrazeny v tabulkách A.7 a A.8.

### 5.1.3 Recenze restaurací – čeština

Pro český dataset recenzí restaurací se nejlepší modely oproti předešlým datasetům mění, protože modely, které byly nejlepší na anglických datech, nemají vícejazyčné nebo české verze. Z vícejazyčných modelů měl nejlepší výsledky model typu XLM-RoBERTa, a to jak ve verzi *base*, tak *large*. U českých modelů byl neúspěšnější FERNET-C5 (model typu BERT), který dosáhl stejné testovací přesnosti jako *large* verze modelu XLM-RoBERTa. U vícejazyčných modelů měla metoda *QA-M* nejlepší výsledky, zatímco u českého modelu byla lepší klasická metoda. Nejlepší modely jsou opět zobrazeny v menší tabulce 5.3. Zajímavé je, že z českých modelů měly lepší výsledky modely založené na BERTu než modely na architektuře RoBERTa, který by měl být vylepšenou verzí modelu BERT.

model	přesnost	v. F1	m. F1
xlm-roberta-base	<b>74,16</b>	<b>74,00</b>	<b>73,15</b>
<b>xlm-roberta-baseQA</b>	<b>78,12</b>	<b>78,21</b>	<b>77,39</b>
xlm-roberta-large	<b>79,33</b>	<b>79,19</b>	<b>78,20</b>
<b>xlm-roberta-largeQA</b>	<b>79,64</b>	<b>79,74</b>	<b>78,87</b>
Czert-B-base-cased	<b>77,81</b>	<b>77,70</b>	<b>76,95</b>
Czert-B-base-casedQA	<b>77,51</b>	<b>77,12</b>	75,98
<b>FERNET-C5</b>	<b>79,64</b>	<b>79,25</b>	<b>78,43</b>
FERNET-C5QA	<b>79,03</b>	<b>78,95</b>	<b>78,45</b>

■ **Tabulka 5.3** Tabulka nejlepších modelů pro dataset Recenze restaurací v češtině ukazující vyhodnocovací metriky na testovacích datech

Z podrobnější tabulky A.9 je patrné, že vícejazyčné modely vykazují lepší výsledky než čistě anglické modely, které se nedokázaly efektivně vypořádat s českými daty. Zajímavou anomálií je model DeBERTa, který při použití metody *QA-M* dosáhl na testovacích datech přesnosti 74,16 %, čímž překonal vícejazyčný model BERT (mBERT) i model trénovaný na slovanských jazycích, SlavicBERT. Vícejazyčné modely měly testovací přesnost v rozmezí 72,64 % až 79,64 %, zatímco čistě anglické modely vykazovaly výsledky mezi 63,83 % a 74,16 %. České modely měly testovací přesnost mezi 74,77 % a 79,64 %, což je podobné výsledkům vícejazyčných modelů. Model SlavicBERT, trénovaný na více slovanských jazycích včetně češtiny, zaostával za čistě českými modely i vícejazyčnými modely. Podrobné výsledky českých modelů a SlavicBERTu jsou uvedeny v tabulce A.10.

V praktických testech byly opět nejlepší modely trénované na češtině a vícejazyčné modely. Celkem 5 modelů mělo všechny odpovědi správně, včetně modelů *FERNET-C5* a *xlm-roberta-base*, které vykazaly nejlepší testovací přesnost (a i F1 skóre). Jen *large* verze modelu XLM-RoBERTa, která měla nejlepší testovací skóre, neprošla plným počtem bodů, protože na test *R2* odpověděl špatně. Ale *large* verze XML-RoBERTa trénovaná klasickou metodou 12 bodů

dosáhla. Z českých modelů dosáhly 12 bodů také modely typu FERNET-C5-RoBERTa. Všechny ostatní české modely měly minimálně 9 správných odpovědí. Model DeBERTa, který měl poměrně vysokou testovací přesnost, měl 10 správných odpovědí, čímž konkuroval i českým modelům. Výsledky všech modelů na praktických testech jsou zobrazeny v tabulkách A.11 a A.12.

#### 5.1.4 Mediální články – čeština

V datasetu mediálních článků byly opět nejlepší výsledky u vícejazyčných a českých modelů. Vícejazyčný model XLM-RoBERTa dosáhl v obou velikostech *base* a *large* nejlepších výsledků, a to při použití klasické metody. Z českých modelů měl nejlepší výsledky FERNET-C5-RoBERTa s metodou *QA-M*, který dokonce překonal obě varianty XLM-RoBERTa. Další modely, které dosáhly dobrých výsledků, jsou zobrazeny v tabulce 5.4. Model FERNET-C5, který byl úspěšný v předchozím datasetu, dosáhl také dobrých výsledků.

model	přesnost	v. F1	m. F1
<b>xlm-roberta-base</b>	<b><u>92,25</u></b>	<b><u>92,29</u></b>	<b><u>92,28</u></b>
xlm-roberta-baseQA	87,45	87,43	87,45
<b>xlm-roberta-large</b>	<b><u>94,10</u></b>	<b><u>94,08</u></b>	<b><u>94,08</u></b>
xlm-roberta-largeQA	<b><u>93,73</u></b>	<b><u>93,74</u></b>	<b><u>93,75</u></b>
robeczech-base	<b><u>91,88</u></b>	<b><u>91,82</u></b>	<b><u>91,82</u></b>
robeczech-baseQA	90,77	90,70	90,76
FERNET-C5	<b><u>92,62</u></b>	<b><u>92,59</u></b>	<b><u>92,61</u></b>
FERNET-C5QA	<b><u>92,25</u></b>	<b><u>92,27</u></b>	<b><u>92,30</u></b>
FERNET-C5-RoBERTa	91,51	91,49	91,47
<b>FERNET-C5-RoBERTaQA</b>	<b><u>94,83</u></b>	<b><u>94,83</u></b>	<b><u>94,83</u></b>

**Tabulka 5.4** Tabulka nejlepších modelů pro dataset Mediální články v češtině ukazující vyhodnocovací metriky na testovacích datech

Z podrobnější tabulky A.13 je patrný velký rozdíl mezi anglickými a vícejazyčnými modely. Vícejazykové modely měly testovací přesnost v rozmezí mezi 85,24 % a 94,10 %, zatímco anglické modely dosahovaly přesnosti pouze mezi 31,73 % a 79,34 %. Tyto výrazné rozdíly jsou způsobeny tím, že dataset mediálních článků obsahuje delší texty než předchozí dataset, což vedlo k problémům u některých anglických modelů, které se během 10 epoch trénování nedokázaly naučit správně rozpoznávat sentiment v jiném jazyce. Další problém spočíval v omezené maximální délce tokenizovaného textu na 256 tokenů, jak bylo uvedeno v sekci 4.1.4, přičemž většina anglických tokenizátorů měla tokenizované texty delší. To mohlo omezit schopnost modelů rozpoznat důležité aspekty v textu. Výsledkem bylo, že některé modely pouze tipovaly sentiment, což vedlo k přesnosti kolem 33 %, což odpovídá náhodnému tipování mezi třemi možnostmi.

České modely (a SlavicBERT) vykazovaly velmi podobné výsledky, přičemž jejich testovací přesnost se pohybovala mezi 88,93 % a 94,83 %. I ostatní metricky F1 skóre byly velmi podobné hodnotám přesnosti, protože tento dataset je vyvážený, jak ukazuje graf 4.1. České modely neměly problém s délkou textu, protože byly schopny efektivně tokenizovat české texty tak, aby nepřekročily hranici 256 tokenů. Výsledky trénování českých modelů (a SlavicBERT) jsou uvedeny v tabulce A.14.

Praktické testy pro tuto doménu byly zajímavé, protože pouze jednomu modelu se podařilo správně predikovat všechny odpovědi ve všech testech. Tím modelem byl *robeczech-base*, který měl testovací přesnost (a obě F1 skóre) o 3 % nižší než nejlepší model. Ostatní české modely v tabulce 5.4 měly správně 11 odpovědí. Všechny chyby se vyskytly ve stejném testu, *L1*, který testoval část věty: „Výkon je dobrý. . .“ a ptal se na aspekt *výkon*. Vícejazyčné modely měly rovněž problémy s praktickými testy, nejlepší model, který správně odpověděl na 11 z 12 testů, byl *xlm-roberta-large*, který rovněž chyboval v testu *L1*. Z tabulek A.15 a A.16 je zřejmé, že české a vícejazyčné modely měly problémy s testovou sadou L, která testovala recenze notebooků.

## 5.2 Porovnání použitých modelů a metod

V předchozích sekcích byly uvedeny výsledky trénování na jednotlivých datasetech, přičemž pro každý dataset byl vybrán nejlepší model, který dosáhl nejlepších výsledků v dané doméně. Pro lepší přehlednost a výběr modelu, který by mohl zvládnout více domén, je nezbytné porovnat výsledky těchto modelů napříč různými datasety. Každý použitý dataset totiž obsahuje některé společné rysy s jinými, ať už jde o jazyk nebo tematické zaměření. Cílem této sekce je porovnat nejlepší modely pro jednotlivé datasety, a to nejen mezi sebou, ale i ve vztahu k použitým metodám trénování. Zároveň bude zhodnoceno, zda některá metoda (klasická nebo *QA-M*) prokazuje výrazně lepší výsledky v určitých podmínkách.

### 5.2.1 Porovnání metod – Klasická vs. QA-M

Jak již bylo zmíněno, každý model, který byl natrénován na všech datasetech, byl trénován pomocí dvou metod. Jedna metoda poskytovala jako vstup jednu větu, zatímco druhá používala dvojici vět, přičemž jedna z nich byla otázkou týkající se sentimentu daného aspektu. Podrobnější popis těchto a dalších nepoužitých metod je uveden v sekci 4.2. Tato podsekce se zaměřuje na porovnání těchto dvou metod napříč všemi trénováními a hodnotí, zda některá z metod vykazuje lepší výsledky. K porovnání jsou využity všechny tabulky v příloze A.

Celkově bylo natrénováno 56 modelů, přičemž každý z nich byl trénován oběma metodami, dohromady teda bylo vytvořeno 112 modelů. Z těchto 56 dvojic modelů dosáhl model trénovaný metodou *QA-M* lepší testovací přesnosti ve 31 případech než model trénovaný klasickou metodou. Také na praktických



testech měly modely trénované metodou *QA-M* tendenci vykazovat lepší výsledky, celkově byly 18krát lepší a 15krát horší než modely trénované klasickou metodou. Často však docházelo k tomu, že model trénovaný metodou *QA-M* měl sice lepší testovací přesnost, ale horší výsledky na praktických testech než stejný model trénovaný klasickou metodou.

Rozdíly mezi těmito metodami nebyly výrazné a v mnoha případech byly výsledky téměř identické. Největší rozdíl, který byl zaznamenán, činil 8 % u modelu *xlm-roberta-base* na datasetu Recenze restaurací v angličtině. U tohoto modelu byl rovněž největší rozdíl v praktických testech, kde model trénovaný metodou *QA-M* vykázal o 5 správných predikcí více. Pokud by model trénovaný klasickou metodou byl znova trénován, mohl by si zlepšit své výsledky, protože u předchozího datasetu měl tento model s oběma metodami velmi podobné výsledky.

I když metoda *QA-M* dosahovala lepších výsledků než klasická metoda, rozdíly nebyly dostatečně velké na to, aby metodě byla přičítána zásadní role. Dokonce bylo více modelů, které dosáhly nejlepších výsledků pro jednotlivé datasety, trénováno klasickou metodou. Z tohoto důvodu se ukazuje, že metoda nehraje tak významnou roli, jako výběr modelu.

### 5.2.2 Porovnání modelů na anglických datech

Na anglických datech dosahovaly nejlepší výsledky novější modely. U obou datasetů byl model ModernBERT ve verzi *large* vždy jedním z nejlepších modelů. Modely DeBERTa a RoBERTa měly rovněž velmi dobré výsledky, přičemž RoBERTa byl obzvláště úspěšný na datasetu Recenze restaurací, kde dokonce překonal i *large* verzi modelu ModernBERT. Oba modely, DeBERTa a RoBERTa, mají také verzi *large*, avšak vzhledem k tomu, že hlavním zaměřením této práce jsou data v češtině, nebyly tyto verze při trénování použity.

Vícejazyčné modely, jako mBERT a XLM-RoBERTa, vykazovaly nižší výkon ve srovnání se svými čistě anglickými alternativami. U datasetu Recenze notebooků měl XLM-RoBERTa sice podobné výsledky jako RoBERTa, ale u druhého datasetu již neprokázal tak dobrý výkon. Tento výsledek je očekávaný, protože modely, které jsou trénovány na více jazycích, obvykle nezvládnou jednotlivý jazyk lépe než modely specializované na tento jazyk.

České modely (a SlavicBERT) měly na anglických datech špatné výsledky, což bylo očekávané. Modely, které byly trénovány na jednom jazyce a následně použity pro analýzu textů v jiném jazyce, obvykle vykazují horší výkon, protože jejich tokenizátor je specializován na daný jazyk. Rozdíly v délce tokenizovaných textů mohou být značné, což ukazují grafy 4.3 a 4.5.

Pro analýzu sentimentu v angličtině je tedy nezbytné používat modely, které byly trénovány na anglických datech, a ideálně by měly být specializované pouze na angličtinu pro lepší přesnost. Na základě pozorovaných výsledků je obtížné doporučit pouze jeden model, protože všechny vylepšené verze modelu BERT dosahovaly podobně dobrých výsledků. Doporučuje se tedy vyzkoušet

všechny tři modely – RoBERTa, DeBERTa a ModernBERT – a testovat, jak se každý z nich chová na různých datasetech.

### 5.2.3 Porovnání modelů na různých jazycích

Díky tomu, že pro češtinu i angličtinu existuje dataset recenzí restaurací, je možné porovnat výkonnost modelů v těchto dvou jazycích. Očekává se, že nejlepší modely pro tyto dva jazyky budou rozdílné, nicméně cílem je spíše zjistit, jak dobře si modely vedou v různých jazycích, a to včetně vícejazyčných modelů, které jsou testovány na obou jazycích.

Vícejazyčné modely, konkrétně XLM-RoBERTa, vykazovaly podobné výsledky u obou jazyků. U anglického datasetu měl sice větší testovací přesnost, ale nižší macro F1 než u českého datasetu. Tento rozdíl je způsoben tím, že český dataset byl více vyvážený než anglický (viz graf 4.1). Zajímavé je, že u obou datasetů měl model *xlm-roberta-base*, trénovaný klasickou metodou, podstatně horší výsledky než stejný model trénovaný metodou *QA-M*. To bylo vidět i na praktických testech, kde modely XLM-RoBERTa vykazovaly podobné výsledky v obou jazycích, přičemž opět *base* verze trénovaná klasickou metodou měla mnohem méně správných odpovědí. Modely mBERT vykazovaly v obou datasetech horší výsledky než XLM-RoBERTa, přičemž tento rozdíl byl větší u českého datasetu.

Zatímco české modely měly podobné výsledky jako XLM-RoBERTa, čistě anglické modely na anglickém datasetu dosahovaly výrazně lepších výsledků než vícejazyčný model, přibližně o 5 %. České modely totiž nevyužívají novější modely, jako jsou DeBERTa a ModernBERT, a proto nemohly dosáhnout takových výsledků, aby překonaly XLM-RoBERTa na českých datech. Pro zlepšení výsledků pro český jazyk je nutné tyto nové modely trénovat čistě na českých datech, čímž by bylo možné využít nové architektury a dosáhnout tak lepších výsledků.

Toto porovnání ukazuje, že i vícejazyčné modely, obvlášt XLM-RoBERTa, dosahují dostatečných výsledků, které v češtině dokonce konkurují modelům trénovaným pouze na českých datech. V angličtině sice vícejazyčný model není dostatečně silný, ale pokud vzniknou vícejazyčné verze novějších modelů (mDeBERTa již existuje, ale není podporováno pro češtinu), mohly by tyto nové vícejazykové modely konkurovat i těm nejlepším modelům trénovaným pouze na anglických datech.

### 5.2.4 Porovnání modelů na českých datech

Na českých datasetech dosahovaly nejlepších výsledků české a vícejazyčné modely. Jak už bylo uvedeno v předchozí podsekci, rozdíl mezi nimi však není velký. Nejčastěji se našel alespoň jeden čistě český model – zejména některá varianta modelů FERNET – jenž překonal *base* verzi XLM-RoBERTa a výkonem se vyrovnal i variantě *large*. Vzhledem k tomu, že české modely dis-

ponují přibližně čtvrtinovým počtem parametrů oproti XLM-RoBERTa, jde o výrazný úspěch. SlavicBERT podával stabilně solidní výkony, nikdy však nepřesáhl výsledky modelů FERNET.

Naopak čistě anglické modely nebyly na českých datech úspěšné, zejména u delších textů v datasetu Mediální texty v češtině. Některé z nich nedokázaly zachytit ani základní strukturu a dosáhly přesnosti kolem 33 %, což odpovídá náhodnému tipování. Důvody tohoto chování již byly rozebrány v části věnované výsledkům na mediálních textech (viz sekce 5.1.4). Pro češtinu tedy anglické modely nejsou vhodným řešením a je nezbytné volit modely, které se dokážou vyrovnat s českými specifiky.

Lze proto doporučit, aby se pro česká data používaly modely přímo trénované na češtině. Přestože rozdíl mezi modelem XLM-RoBERTa a nejlepšími českými modely nebyl dramatický, v praxi je potřeba vyzkoušet více kandidátů a zvolit ten, který se na konkrétní doméně chová nejlépe. V obou českých datasetech si nejlépe vedly modely FERNET, jež v některých případech dosáhla stejných či dokonce lepších výsledků než *large* varianta modelu XLM-RoBERTa. Ani další české modely – Czert nebo RobeCzech – však nepropadly, a zaslouží si podrobnější porovnání při aplikaci na jiné korpusy.

## 5.3 Shrnutí a doporučení

V rámci této práce bylo natrénováno celkem 112 modelů, které se lišily architekturou, trénovací metodou i datovou sadou. Byly zkoumány *anglické*, *české* a *vícejazyčné* varianty, což umožnilo sledovat, jak si modely zaměřené na jeden jazyk vedou na datech v jiných jazycích a naopak. Kvůli limitacím trénovacího prostředí byl každý model natrénován pouze jednou a s jedinou sadou hyperparametrů. Pro důkladnější srovnání by bylo vhodné každý model opakovaně trénovat – jak se stejnými, tak s různými hyperparametry – a ověřit, zda některé modely v průměru překonávají ostatní a o kolik.

Níže je uvedeno stručné zhodnocení dosažených výsledků a doporučení pro praktické nasazení v českých mediálních textech. Dále jsou porovnány vybrané velké jazykové modely (LLM) bez dodatečného přeučování, využívající jediný prompt k analýze sentimentu.

### 5.3.1 Anglické datasety

Nejlepší výsledky opakovaně vykazovaly čistě anglické modely z nejnovějších architektur – ModernBERT, RoBERTa a DeBERTa.

- Ve variantě *large* dosáhl ModernBERT nejvyšší přesnosti na obou doménách (recenze notebooků, recenze restaurací).
- Ve variantě *base* byly rozdíly mezi trojicí ModernBERT / RoBERTa / DeBERTa nepatrné; vítěz se lišil podle domény.

- Volba trénovací metody (**klasická** vs. **QA-M**) výsledky významně neovlivnila, někdy byla lepší klasická a někdy QA-M metoda.

### 5.3.2 České datasety

U českých korpusů se výkony českých a vícejazyčných modelů (hlavně XLM-RoBERTa, mBERT zaostával) velmi přibližovaly.

- Modely **FERNET** (BERT i RoBERTa) konzistentně překonala *base* verzi XLM-RoBERTa a v některých případech se dotáhla i na její *large* variantu, přestože má čtvrtinový počet parametrů.
- Stejně jako v angličtině, ani zde nehrála trénovací metoda zásadní roli; rozdíly byly nepatrné.

### 5.3.3 Doporučení pro české mediální texty

1. **Preferovat čistě české modely.** Zaměřený tokenizátor lépe zvládá diakritiku a morfologii a efektivně tokenizuje delší texty.
2. **Vyzkoušet zejména modely FERNET.** Varianta *base* poskytuje výkon srovnatelný s mnohonásobně větší *XLM-RoBERTa-large*, a přitom je podstatně úspornější.
3. **Dlouhé články segmentovat.** Všechny české modely akceptují maximálně 512 tokenů; dlouhé mediální texty je nutné rozdělit do kratších pasáží, aby nebyly oříznuty.
4. **Metodu trénování volit podle možností.** Rozdíly mezi klasickou a QA-M metodou byly nepatrné, a tak lze preferovat jednodušší implementaci.
5. **Sledovat vývoj nových architektur.** ModernBERT nabízí kontext 8192 tokenů, ale zatím není k dispozici v češtině; česká adaptace podobně „dlouhých“ encoderů by mohla dále zvýšit přesnost v mediální doméně.

### 5.3.4 Analýza sentimentu pomocí LLM

V dnešní době jsou velké jazykové modely (*Large language models, LLM*) rozšířené a veřejně dostupné na internetu. Zatímco *base* verze modelů použitých v této práci mají 110 – 270 milionů parametrů, i nejmenší LLM zpravidla přesahují jednu miliardu parametrů a největší systémy, jako GPT-4 nebo Gemini Ultra, se pravděpodobně pohybují v řádu bilionů parametrů (angl. trillions). Tato čísla jsou však pouze odhady, protože většina společností jejich přesnou velikost neuvádí. Poslední oficiálně zveřejněný údaj se týkal modelu GPT-3, který má 178 miliard parametrů, tedy zhruba tisíckrát více, než má většina modelů použitých v této práci. [142]

#### 5.3.4.1 Popis modelů a použitého promptu

K porovnání bylo vybráno šest veřejně a zdarma dostupných LLM: ChatGPT 4o mini [143], Gemma 3 1B [144], Qwen 2.5 Plus [145], DeepSeek [146], Claude 3.7 Sonnet [147] a Phi-4 [148]. Záměrem bylo zvolit co nejmenší variantu, aby bylo srovnání s modely z této práce co nejpřesnější. Protože většina nástrojů počet parametrů nezveřejňuje, není možné spolehlivě ověřit, zda jde skutečně o nejmenší verzi.

K vygenerování výsledků byl použit jediný prompt, který nejprve popsal úkol a následně předal aspekty ve formátu „aspekt, aspekt“ a text jako poslední informaci. Modely tedy hodnotily všechny aspekty v jedné odpovědi. Prompt byl česky pro české testy (ukázka 5.1) a anglicky pro anglické testy (ukázka 5.2).

```
f"""Prosím proved' aspektovou analýzu sentimentu na níže vložený
text. Tvůj úkol je v textu detekovat předem dané aspekty a pak
následně k nim přiřadit správný sentiment. Nijak tvoje výsledky
nekomentuj, jen ke každému aspektu vrať jeden ze tří sentimentů
(negativní, neutrální a pozitivní).
Aspekty: ..., ...
Text: ..."""
```

■ **Výpis kódu 5.1** Ukázka promptu v češtině pro analýzu sentimentu u LLM, vlastní práce

```
f"""Please perform an aspect sentiment analysis on the text
below. Your task is to detect predetermined aspects in the text
and then assign the correct sentiment to them. Do not comment
on your results, just return one of the three sentiments
(negative, neutral, and positive) for each aspect.
Aspects: ..., ...
Text: ..."""
```

■ **Výpis kódu 5.2** Ukázka promptu v angličtině pro analýzu sentimentu u LLM, vlastní práce

#### 5.3.4.2 Výsledky a závěr k LLM

Všechny modely úkol zvládly bez větších potíží. Jedinou výjimkou byl model Gemma 3 1B, který v anglických testech chybovala v testu *M1* (neutrální sentiment) a v českých datech udělal tři chyby v testech *L1*, *M1* a *O1*. Ostatní modely odpověděly ve všech případech správně, proto k nim není sestavena samostatná tabulka výsledků.

Z toho plyne, že vybraná LLM zvládají aspektovou analýzu sentimentu spolehlivě, avšak jejich podstatně větší velikost (i nejmenší z nich má téměř desetkrát více parametrů než modely použité v této práci) je činí nepraktickými pro analýzu rozsáhlých datových souborů: delší doba predikce a vyšší výpočetní nároky znamenají vyšší provozní náklady oproti menším modelům.

## 5.4 Další výzkum

V této práci byla zkoumána pouze jedna dílčí úloha ABSA, konkrétně *Aspect Sentiment Classification* (ASC), která předpokládá, že sledovaný aspekt je předem známý a pro něj se následně predikuje sentiment. V praxi však tato informace nemusí být dostupná, a modely natrénované v této práci by v takovém případě nebyly schopny sentiment určit. Z tohoto důvodu se nabízí složené úlohy ABSA, například *E2E-ABSA*, jež ke klasifikaci sentimentu nejprve z textu extrahuje aspekty, a lze ji tedy použít i na dosud neanalyzované texty. Další významnou úlohou je *TASD*, kde je kromě sentimentu k aspektu predikována také kategorie, která aspekt přesněji charakterizuje.

Limitem této práce je rovněž to, že každý model byl natrénován pouze jednou. Pro spolehlivější porovnání by bylo vhodné trénování opakovat s více náhodnými inicializacemi i různými sadami hyperparametrů.

Budoucí výzkum se může soustředit právě na složené úlohy *E2E-ABSA* a *TASD*, aby bylo možné přesněji modelovat sentiment v situacích, kdy aspekty nejsou předem definovány. Kromě toho je žádoucí sledovat vývoj encoderových architektur, například odvozenin BERTu (jako je DeBERTa a ModernBERT), a soustavně trénovat nové modely na českých datech, čímž by se rozšířily zdroje pro další výzkum a praktické aplikace.

## Závěr

Tato bakalářská práce se soustředila na aspektovou analýzu sentimentu a přispěla k jejímu rozvoji zejména v českém jazyce a mediální doméně. Úvodní kapitola představila základní principy zpracování přirozeného jazyka a popsala klíčové jazykové modely, které tvoří teoretické východisko pro všechny pozdější části práce.

Následující kapitola shrnula historický vývoj oboru, ukázala posun od statistických přístupů k moderním transformátorovým architekturám a podrobně popsala jednotlivé podúlohy ABSA i navazující složené úlohy. Samostatná třetí kapitola rozšířila perspektivu o český jazyk, zmapovala stávající výzkum a poukázala na specifické výzvy vyplývající z menší dostupnosti kvalitních datasetů.

Čtvrtá kapitola podrobně popsala veškerá použitá data, metody a modely. Pozornost byla věnována i postupu trénování jednotlivých modelů, definici vyhodnocovacích metrik a návrhu praktických testů. Po rozsáhlém experimentálním srovnání několika desítek modelů napříč doménami byla identifikována řešení, která poskytují nejlepší výsledky pro jednotlivé jazyky a domény. Tyto poznatky byly syntetizovány do souboru doporučení, který usnadní volbu vhodného modelu pro české mediální texty.

Z hlediska naplnění cílů bylo vytvořeno komplexní shrnutí dosavadních přístupů k analýze sentimentu a dílčím úlohám ABSA, provedeno experimentální srovnání vybraných modelů napříč jazyky a doménami a – na základě těchto výsledků – doporučeny konkrétní postupy a modely pro praktickou práci s českými texty. Tím práce splnila všechny vytyčené cíle a poskytla podklad pro další výzkum i nasazení ABSA v českém jazykovém prostředí.

## Tabulky výsledků trénování a praktických testů

Tato příloha obsahuje všechny výsledky trénování modelů a jejich následného testování, jak bylo popsáno v předchozích kapitolách. Všechny modely, konkrétně čtrnáct popsaných v sekci 4.3, byly použity pro trénování. Každý model byl natrénován dvěma metodami, jak je uvedeno v sekci 4.2. Celkově tedy bylo natrénováno 28 modelů pro každý dataset uvedený v sekci 4.1, což dohromady představuje 112 natrénovaných modelů.

Pro každý dataset jsou v příloze uvedeny čtyři tabulky – dvě pro trénovací výsledky a dvě pro výsledky praktických testů. Tabulky jsou rozděleny podle typů modelů: Inovativní modely a Jazykově adaptované modely. Přípona *QA* u jména modelu značí použití metody *QA-M*.

**Trénovací tabulky** Trénovací tabulky obsahují detailní informace o procesu trénování jednotlivých modelů. V těchto tabulkách je uvedena doba trénování každého modelu, stejně jako epocha, která vykazala nejlepší výsledky. Dále je uvedena přesnost na validační množině (*val. přesnost*), podle které byl vybrán nejlepší model během samotného trénování. U testovacích dat jsou pak uvedeny všechny metriky popsané v sekci 4.5.1 a to přesnost (*test. přesnost*), vážená F1 (*v. F1*) a macro F1 (*m. F1*). V každé tabulce jsou zvýrazněny nejlepší výsledky – tučně jsou vyznačeny čtyři nejlepší hodnoty pro každou metriku. Nejlepší výsledky jsou pak ještě podržené. Jméno nejlepšího modelu je pak zvýrazněno tučně – u Inovativních modelů jsou vybrány dva, jeden ve velikosti *base* a druhý ve velikosti *large*.

**Tabulky praktických testů** Tabulky praktických testů zobrazují výsledky z praktických testů, které jsou uvedeny v sekci 4.5.2. Modely s nejlepšími výsledky jsou v těchto tabulkách zvýrazněny.



model	trénování	nej. epocha	val. přesnost	test. přesnost	v. F1	m. F1
bert-base-uncased	13m32s	7.	76,19	79,31	79,12	77,47
bert-base-uncasedQA	13m32s	10.	77,70	79,74	79,45	78,10
bert-large-uncased	44m19s	9.	78,14	80,17	79,68	78,29
bert-large-uncasedQA	50m22s	4.	78,79	80,60	80,36	78,62
bert-base-multilingual-cased	15m20s	7.	73,38	75,43	75,19	74,03
bert-base-multilingual-casedQA	15m37s	8.	74,46	76,29	75,95	74,30
roberta-base	14m04s	9.	80,30	80,17	79,44	77,64
roberta-baseQA	17m18s	10.	82,25	81,47	81,10	79,56
xlm-roberta-base	18m19s	8.	78,79	81,03	80,69	78,81
xlm-roberta-baseQA	24m07s	5.	77,92	81,47	80,97	79,36
xlm-roberta-large	54m02s	9.	79,23	81,03	80,32	78,43
xlm-roberta-largeQA	57m55s	9.	79,65	79,74	79,15	77,09
deberta-v3-base	24m31s	10.	<b>83,12</b>	<b>82,76</b>	<b>82,29</b>	<b>80,60</b>
deberta-v3-baseQA	19m25s	9.	<u><b>83,55</b></u>	81,90	81,67	<b>80,31</b>
ModernBERT-base	24m04s	7.	81,82	81,47	81,04	79,23
<b>ModernBERT-baseQA</b>	19m23s	8.	82,46	<u><b>84,48</b></u>	<u><b>84,19</b></u>	<u><b>82,65</b></u>
<b>ModernBERT-large</b>	54m38s	8.	<b>83,12</b>	<u><b>84,91</b></u>	<u><b>84,49</b></u>	<u><b>82,97</b></u>
ModernBERT-largeQA	55m24s	7.	<u><b>83,55</b></u>	<b>82,33</b>	<b>81,79</b>	79,86

■ **Tabulka A.1** Inovativní modely – trénování na datasetu Recenze notebooků v angličtině

model	trénování	nej. epocha	val. přesnost	test. přesnost	v. F1	m. F1
bert-base-bg-cs-pl-ru-cased	17m49s	4.	70,13	64,66	64,57	62,03
bert-base-bg-cs-pl-ru-casedQA	21m42s	9.	<b>70,35</b>	<b>71,55</b>	<b>71,43</b>	<b>69,62</b>
Czert-B-base-cased	13m21s	10.	<b>71,21</b>	66,38	65,94	64,51
Czert-B-base-casedQA	13m53s	10.	68,61	70,26	69,79	67,72
robeczech-base	13m55s	7.	69,70	70,26	69,82	68,48
robeczech-baseQA	13m38s	9.	<u><b>72,51</b></u>	<b>70,69</b>	<b>70,20</b>	<b>68,84</b>
<b>FERNET-C5</b>	15m22s	7.	69,26	<u><b>72,84</b></u>	<u><b>72,25</b></u>	<u><b>70,41</b></u>
FERNET-C5QA	14m53s	9.	69,05	<b>71,12</b>	<b>70,62</b>	<b>68,52</b>
FERNET-C5-RoBERTa	14m34s	7.	68,61	65,09	64,16	62,17
FERNET-C5-RoBERTaQA	14m31s	9.	<b>72,08</b>	<b>70,69</b>	69,57	67,11

■ **Tabulka A.2** Jazykově adaptované modely – trénování na datasetu Recenze notebooků v angličtině

model	L1	L2	L3	R1	R2	R3	R4	M1	M2	M3	O1	O2	skóre
bert-base-uncased	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	11/12
bert-base-uncasedQA	✓	✓	✓	✓	✗	✓	✓	✗	✓	✗	✓	✓	09/12
bert-large-uncased	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	11/12
bert-large-uncasedQA	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	11/12
bert-base-multilingual-cased	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✗	07/12
bert-base-multilingual-casedQA	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✗	✓	09/12
roberta-base	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	11/12
<b>roberta-baseQA</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	<b>12/12</b>
xlm-roberta-base	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	10/12
xlm-roberta-baseQA	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	10/12
xlm-roberta-large	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	11/12
xlm-roberta-largeQA	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✗	✓	09/12
deberta-v3-base	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	11/12
deberta-v3-baseQA	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	11/12
ModernBERT-base	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	11/12
ModernBERT-baseQA	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	11/12
ModernBERT-large	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	11/12
<b>ModernBERT-largeQA</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	<b>12/12</b>

■ **Tabulka A.3** Inovativní modely – praktické testy pro dataset Recenze notebooků v angličtině

model	L1	L2	L3	R1	R2	R3	R4	M1	M2	M3	O1	O2	skóre
bert-base-bg-cs-pl-ru-cased	✓	✗	✓	✓	✓	✗	✗	✗	✗	✓	✗	✓	06/12
bert-base-bg-cs-pl-ru-casedQA	✓	✗	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	08/12
<b>Czert-B-base-cased</b>	✓	✗	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	<b>09/12</b>
Czert-B-base-casedQA	✓	✗	✓	✓	✗	✓	✓	✗	✗	✓	✓	✓	08/12
robeczech-base	✓	✗	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	07/12
<b>robeczech-baseQA</b>	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✗	<b>09/12</b>
FERNET-C5	✓	✗	✓	✓	✓	✗	✗	✗	✗	✓	✓	✗	06/12
FERNET-C5QA	✓	✓	✗	✓	✓	✗	✗	✗	✓	✓	✓	✗	07/12
FERNET-C5-RoBERTa	✓	✗	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	08/12
FERNET-C5-RoBERTaQA	✓	✗	✓	✗	✗	✓	✓	✗	✗	✓	✓	✗	06/12

■ **Tabulka A.4** Jazykově adaptované modely – praktické testy pro dataset Recenze notebooků v angličtině

model	trénování	nej. epocha	val. přesnost	test. přesnost	v. F1	m. F1
bert-base-uncased	20m34s	7.	82,78	80,33	80,27	74,36
bert-base-uncasedQA	20m46s	10.	83,33	81,99	82,21	77,22
bert-large-uncased	1h15m14s	10.	84,17	84,76	84,80	80,13
bert-large-uncasedQA	1h08m51s	6.	85,00	83,93	83,77	79,25
bert-base-multilingual-cased	22m37s	5.	80,42	78,67	78,19	72,67
bert-base-multilingual-casedQA	23m11s	6.	79,58	80,33	80,47	75,36
<b>roberta-base</b>	21m16s	8.	<b><u>86,94</u></b>	<b><u>86,70</u></b>	<b><u>86,72</u></b>	<b><u>83,02</u></b>
roberta-baseQA	21m09s	7.	85,69	85,04	84,83	79,94
xlm-roberta-base	25m51s	6.	78,06	74,52	72,00	62,77
xlm-roberta-baseQA	30m42s	6.	86,11	82,55	82,01	76,24
xlm-roberta-large	1h19m36s	10.	85,42	81,99	81,99	77,19
xlm-roberta-largeQA	1h15m33s	10.	85,97	80,06	80,07	73,65
deberta-v3-base	28m54s	10.	85,00	<b>85,60</b>	<b>85,74</b>	<b>81,51</b>
deberta-v3-baseQA	28m02s	9.	85,00	84,21	84,51	79,57
ModernBERT-base	29m09s	9.	<b><u>86,94</u></b>	83,66	83,72	78,38
ModernBERT-baseQA	31m18s	3.	86,67	82,83	82,37	76,37
<b>ModernBERT-large</b>	1h23m23s	4.	<b><u>89,31</u></b>	<b><u>86,15</u></b>	<b><u>85,74</u></b>	<b><u>81,09</u></b>
ModernBERT-largeQA	1h24m47s	7.	<b>87,92</b>	<b>85,87</b>	<b><u>85,74</u></b>	<b>81,08</b>

■ **Tabulka A.5** Inovativní modely – trénování na datasetu Recenze restaurací v angličtině

model	trénování	nej. epocha	val. přesnost	test. přesnost	v. F1	m. F1
bert-base-bg-cs-pl-ru-cased	23m58s	9.	78,19	74,24	74,64	67,14
bert-base-bg-cs-pl-ru-casedQA	23m28s	9.	78,33	<b>76,73</b>	<b>76,89</b>	<b>70,15</b>
Czert-B-base-cased	21m07s	10.	77,78	<b>77,01</b>	<b>76,80</b>	<b>71,01</b>
Czert-B-base-casedQA	23m26s	7.	77,22	74,24	74,02	67,30
robeczech-base	24m53s	10.	<b>78,89</b>	74,79	74,48	67,46
robeczech-baseQA	21m46s	8.	77,78	<b>77,01</b>	<b>76,55</b>	<b>70,07</b>
FERNET-C5	25m54s	7.	<b>78,47</b>	74,24	74,74	67,71
<b>FERNET-C5QA</b>	21m18s	10.	<b>78,89</b>	<b>77,84</b>	<b>77,90</b>	<b>72,03</b>
FERNET-C5-RoBERTa	20m31s	10.	77,64	74,52	73,83	66,74
FERNET-C5-RoBERTaQA	24m22s	10.	<b>78,75</b>	74,24	73,93	66,64

■ **Tabulka A.6** Jazykově adaptované modely – trénování na datasetu Recenze restaurací v angličtině

model	L1	L2	L3	R1	R2	R3	R4	M1	M2	M3	O1	O2	skóre
<b>bert-base-uncased</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
<b>bert-base-uncasedQA</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
bert-large-uncased	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	11/12
<b>bert-large-uncasedQA</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
bert-base-multilingual-cased	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	10/12
bert-base-multilingual-casedQA	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	10/12
<b>roberta-base</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
<b>roberta-baseQA</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
xlm-roberta-base	✗	✓	✗	✗	✗	✓	✓	✗	✓	✗	✗	✓	05/12
xlm-roberta-baseQA	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	10/12
xlm-roberta-large	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	11/12
xlm-roberta-largeQA	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	10/12
<b>deberta-v3-base</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
<b>deberta-v3-baseQA</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
<b>ModernBERT-base</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
<b>ModernBERT-baseQA</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
<b>ModernBERT-large</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
<b>ModernBERT-largeQA</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12

■ **Tabulka A.7** Inovativní modely – praktické testy pro dataset Recenze restaurací v angličtině

model	L1	L2	L3	R1	R2	R3	R4	M1	M2	M3	O1	O2	skóre
bert-base-bg-cs-pl-ru-cased	✓	✗	✓	✓	✓	✓	✓	✗	✓	✗	✓	✓	09/12
bert-base-bg-cs-pl-ru-casedQA	✓	✓	✗	✓	✓	✓	✓	✗	✓	✗	✓	✓	09/12
Czert-B-base-cased	✓	✓	✓	✗	✗	✓	✓	✓	✗	✓	✓	✓	09/12
<b>Czert-B-base-casedQA</b>	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	<b>11/12</b>
<b>robeczech-base</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	<b>11/12</b>
robeczech-baseQA	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✓	10/12
FERNET-C5	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✓	10/12
FERNET-C5QA	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✓	10/12
FERNET-C5-RoBERTa	✓	✓	✗	✓	✓	✗	✓	✗	✓	✗	✓	✓	08/12
FERNET-C5-RoBERTaQA	✓	✓	✗	✓	✓	✓	✓	✗	✓	✗	✓	✗	08/12

■ **Tabulka A.8** Jazykově adaptované modely – praktické testy pro dataset Recenze restaurací v angličtině



model	trénování	nej. epocha	val. přesnost	test. přesnost	v. F1	m. F1
bert-base-uncased	23m05s	10.	75,15	71,43	71,40	70,83
bert-base-uncasedQA	23m34s	9.	73,93	70,21	70,03	69,12
bert-large-uncased	1h06m05s	8.	73,93	67,48	67,07	66,26
bert-large-uncasedQA	1h06m41s	8.	71,95	67,17	66,81	66,04
bert-base-multilingual-cased	22m22s	10.	77,13	73,25	73,26	72,97
bert-base-multilingual-casedQA	20m34s	10.	76,67	72,64	72,60	71,80
roberta-base	20m23s	8.	72,10	63,83	63,63	62,98
roberta-baseQA	23m10s	6.	71,04	68,69	68,51	68,04
xlm-roberta-base	24m08s	5.	<b>79,42</b>	<b>74,16</b>	<b>74,00</b>	<b>73,15</b>
<b>xlm-roberta-baseQA</b>	26m23s	10.	<b><u>80,34</u></b>	<b><u>78,12</u></b>	<b><u>78,21</u></b>	<b><u>77,39</u></b>
xlm-roberta-large	1h14m09s	8.	<b><u>84,91</u></b>	<b>79,33</b>	<b>79,19</b>	<b>78,20</b>
<b>xlm-roberta-largeQA</b>	1h19m23s	7.	<b><u>84,76</u></b>	<b><u>79,64</u></b>	<b><u>79,74</u></b>	<b><u>78,87</u></b>
deberta-v3-base	27m16s	9.	76,52	71,73	71,37	70,16
deberta-v3-baseQA	25m55s	8.	78,05	<b>74,16</b>	73,84	73,11
ModernBERT-base	27m33s	7.	71,65	65,65	64,79	63,67
ModernBERT-baseQA	26m49s	7.	73,78	65,35	64,68	63,67
ModernBERT-large	1h19m05s	8.	74,24	68,69	68,52	67,80
ModernBERT-largeQA	1h15m43s	9.	73,78	71,43	71,31	70,69

■ **Tabulka A.9** Inovativní modely – trénování na datasetu Recenze restaurací v češtině

model	trénování	nej. epocha	val. přesnost	test. přesnost	v. F1	m. F1
bert-base-bg-cs-pl-ru-cased	22m19s	6.	80,03	73,25	73,19	72,49
bert-base-bg-cs-pl-ru-casedQA	26m36s	9.	79,12	71,43	71,09	70,21
Czert-B-base-cased	20m03s	8.	80,49	<b>77,81</b>	<b>77,70</b>	<b>76,95</b>
Czert-B-base-casedQA	20m04s	3.	80,79	<b>77,51</b>	<b>77,12</b>	75,98
robeczech-base	23m15s	10.	<b>83,08</b>	76,90	76,90	<b>76,14</b>
robeczech-baseQA	19m50s	4.	81,86	74,77	74,77	74,17
<b>FERNET-C5</b>	20m35s	5.	<b><u>83,23</u></b>	<b><u>79,64</u></b>	<b><u>79,25</u></b>	<b><u>78,43</u></b>
FERNET-C5QA	20m21s	6.	<b>82,32</b>	<b>79,03</b>	<b>78,95</b>	<b><u>78,45</u></b>
FERNET-C5-RoBERTa	22m54s	8.	<b>83,23</b>	76,60	76,37	75,38
FERNET-C5-RoBERTaQA	22m31s	8.	<b>82,32</b>	76,90	76,71	75,80

■ **Tabulka A.10** Jazykově adaptované modely – trénování na datasetu Recenze restaurací v češtině

model	L1	L2	L3	R1	R2	R3	R4	M1	M2	M3	O1	O2	skóre
bert-base-uncased	✓	✓	✓	✗	✗	✓	✓	✗	✗	✓	✓	✓	08/12
bert-base-uncasedQA	✓	✗	✓	✗	✗	✓	✓	✗	✗	✓	✓	✗	06/12
bert-large-uncased	✓	✗	✓	✗	✗	✓	✓	✗	✗	✓	✗	✓	06/12
bert-large-uncasedQA	✓	✗	✓	✗	✗	✓	✓	✗	✓	✗	✓	✗	06/12
bert-base-multilingual-cased	✓	✗	✓	✗	✓	✓	✓	✗	✗	✓	✓	✓	08/12
bert-base-multilingual-casedQA	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	10/12
roberta-base	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	09/12
roberta-baseQA	✓	✗	✓	✗	✗	✓	✓	✗	✗	✓	✗	✗	05/12
xlm-roberta-base	✓	✓	✗	✗	✗	✓	✓	✗	✗	✓	✓	✗	06/12
<b>xlm-roberta-baseQA</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	<b>12/12</b>
<b>xlm-roberta-large</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	<b>12/12</b>
xlm-roberta-largeQA	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	11/12
deberta-v3-base	✓	✓	✗	✗	✗	✓	✓	✗	✓	✓	✓	✓	08/12
deberta-v3-baseQA	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	10/12
ModernBERT-base	✓	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗	✓	04/12
ModernBERT-baseQA	✓	✗	✓	✗	✗	✓	✓	✗	✗	✓	✓	✗	06/12
ModernBERT-large	✓	✗	✓	✗	✗	✓	✓	✗	✗	✓	✓	✗	06/12
ModernBERT-largeQA	✓	✗	✗	✗	✗	✓	✓	✓	✗	✗	✓	✓	06/12

■ **Tabulka A.11** Inovativní modely – praktické testy pro dataset Recenze restaurací v češtině

model	L1	L2	L3	R1	R2	R3	R4	M1	M2	M3	O1	O2	skóre
bert-base-bg-cs-pl-ru-cased	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✓	10/12
bert-base-bg-cs-pl-ru-casedQA	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✓	10/12
Czert-B-base-cased	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	11/12
Czert-B-base-casedQA	✓	✓	✗	✓	✓	✓	✓	✗	✗	✓	✓	✓	09/12
robeczech-base	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	11/12
robeczech-baseQA	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	11/12
<b>FERNET-C5</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	<b>12/12</b>
FERNET-C5QA	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	11/12
<b>FERNET-C5-RoBERTa</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	<b>12/12</b>
<b>FERNET-C5-RoBERTaQA</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	<b>12/12</b>

■ **Tabulka A.12** Jazykově adaptované modely – praktické testy pro dataset Recenze restaurací v češtině

model	trénování	nej. epocha	val. přesnost	test. přesnost	v. F1	m. F1
bert-base-uncased	16m08s	3.	39,30	33,95	25,82	26,50
bert-base-uncasedQA	18m58s	7.	41,14	40,59	35,39	35,99
bert-large-uncased	56m27s	2.	35,42	31,73	15,29	16,06
bert-large-uncasedQA	58m04s	2.	38,19	33,58	22,64	23,43
bert-base-multilingual-cased	17m59s	8.	85,24	85,24	85,24	85,30
bert-base-multilingual-casedQA	19m41s	10.	88,38	<b>87,82</b>	<b>87,84</b>	<b>87,80</b>
roberta-base	15m58s	2.	38,93	33,21	22,13	22,91
roberta-baseQA	19m00s	10.	41,88	39,48	36,68	37,15
<b>xlm-roberta-base</b>	21m48s	9.	<b><u>92,62</u></b>	<b><u>92,25</u></b>	<b><u>92,29</u></b>	<b><u>92,28</u></b>
xlm-roberta-baseQA	20m52s	9.	<b>88,56</b>	87,45	87,43	87,45
<b>xlm-roberta-large</b>	1h02m33s	8.	<b><u>93,91</u></b>	<b><u>94,10</u></b>	<b><u>94,08</u></b>	<b><u>94,08</u></b>
xlm-roberta-largeQA	1h05m35s	8.	<b>92,80</b>	<b>93,73</b>	<b>93,74</b>	<b>93,75</b>
deberta-v3-base	24m25s	10.	67,53	61,99	61,19	61,60
deberta-v3-baseQA	23m41s	6.	46,31	43,17	34,08	34,92
ModernBERT-base	25m02s	9.	68,08	72,32	72,17	72,07
ModernBERT-baseQA	23m42s	10.	72,32	75,28	75,30	75,26
ModernBERT-large	1h05m46s	9.	81,37	77,12	77,19	77,16
ModernBERT-largeQA	1h08m46s	10.	79,89	79,34	79,50	79,44

■ **Tabulka A.13** Inovativní modely – trénování na datasetu Mediální texty v češtině

model	trénování	nej. epocha	val. přesnost	test. přesnost	v. F1	m. F1
bert-base-bg-cs-pl-ru-cased	18m22s	10.	89,85	90,77	90,79	90,78
bert-base-bg-cs-pl-ru-casedQA	18m07s	10.	88,93	91,51	91,52	91,52
Czert-B-base-cased	16m25s	9.	89,85	88,93	88,91	88,96
Czert-B-base-casedQA	16m52s	9.	90,41	90,04	90,02	90,09
robeczech-base	16m09s	9.	93,36	<b>91,88</b>	<b>91,82</b>	<b>91,82</b>
robeczech-baseQA	16m34s	8.	<b>93,91</b>	90,77	90,70	90,76
FERNET-C5	18m44s	7.	<b>93,91</b>	<b>92,62</b>	<b>92,59</b>	<b>92,61</b>
FERNET-C5QA	17m54s	6.	<b>93,54</b>	<b>92,25</b>	<b>92,27</b>	<b>92,30</b>
FERNET-C5-RoBERTa	16m03s	8.	93,17	91,51	91,49	91,47
<b>FERNET-C5-RoBERTaQA</b>	16m39s	9.	<u><b>94,10</b></u>	<u><b>94,83</b></u>	<u><b>94,83</b></u>	<u><b>94,83</b></u>

■ **Tabulka A.14** Jazykově adaptované modely – trénování na datasetu Mediální texty v češtině

model	L1	L2	L3	R1	R2	R3	R4	M1	M2	M3	O1	O2	skóre
bert-base-uncased	✓	✗	✓	✗	✗	✓	✓	✗	✓	✗	✓	✗	06/12
bert-base-uncasedQA	✓	✗	✓	✗	✗	✓	✓	✗	✓	✗	✓	✗	06/12
bert-large-uncased	✓	✗	✓	✗	✗	✓	✓	✗	✓	✗	✓	✗	06/12
bert-large-uncasedQA	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✗	✓	07/12
bert-base-multilingual-cased	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	09/12
bert-base-multilingual-casedQA	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	10/12
roberta-base	✓	✗	✓	✗	✗	✓	✓	✗	✓	✗	✓	✗	06/12
roberta-baseQA	✓	✗	✓	✗	✗	✗	✗	✓	✗	✓	✗	✗	04/12
xlm-roberta-base	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	10/12
xlm-roberta-baseQA	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	10/12
<b>xlm-roberta-large</b>	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	<b>11/12</b>
xlm-roberta-largeQA	✗	✓	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	09/12
deberta-v3-base	✗	✗	✓	✓	✗	✗	✗	✓	✓	✓	✓	✗	06/12
deberta-v3-baseQA	✓	✗	✓	✗	✗	✓	✓	✗	✗	✓	✓	✗	06/12
ModernBERT-base	✗	✗	✗	✗	✓	✓	✓	✓	✗	✗	✓	✗	05/12
ModernBERT-baseQA	✗	✗	✓	✗	✗	✓	✓	✓	✓	✗	✓	✓	07/12
ModernBERT-large	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10/12
ModernBERT-largeQA	✗	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗	08/12

■ **Tabulka A.15** Inovativní modely – praktické testy pro dataset Mediální texty v češtině

model	L1	L2	L3	R1	R2	R3	R4	M1	M2	M3	O1	O2	skóre
bert-base-bg-cs-pl-ru-cased	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	09/12
bert-base-bg-cs-pl-ru-casedQA	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11/12
Czert-B-base-cased	✗	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	09/12
Czert-B-base-casedQA	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10/12
<b>robeczech-base</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	<b>12/12</b>
robeczech-baseQA	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11/12
FERNET-C5	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11/12
FERNET-C5QA	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11/12
FERNET-C5-RoBERTa	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11/12
FERNET-C5-RoBERTaQA	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11/12

■ **Tabulka A.16** Jazykově adaptované modely – praktické testy pro dataset Mediální texty v češtině



# Bibliografie

1. KHURANA, Diksha; KOLI, Aditya; KHATTER, Kiran; SINGH, Sukhdev. Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*. 2023, roč. 82, č. 3, s. 3713–3744. ISSN 1573-7721. Dostupné z DOI: 10.1007/s11042-022-13428-4.
2. ÚSTAV PRO JAZYK ČESKÝ AV ČR, v. v. i. *Internetová jazyková příručka* [online]. [cit. 2025-03-29]. Dostupné z: <https://prirucka.ujc.cas.cz/?id=778>. Morfematika.
3. W3TECHS. *Usage statistics of content languages for websites* [online]. 2025. [cit. 2025-03-28]. Dostupné z: [https://w3techs.com/technologies/overview/content\\_language](https://w3techs.com/technologies/overview/content_language).
4. HIEMSTRA, Djoerd. Language Models. In: *Encyclopedia of Database Systems*. Ed. LIU, LING; ÖZSU, M. TAMER. Boston, MA: Springer US, 2009, s. 1591–1594. ISBN 978-0-387-39940-9. Dostupné z DOI: 10.1007/978-0-387-39940-9\_923.
5. WEI, Chengwei; WANG, Yun-Cheng; WANG, Bin; KUO, C.-C. Jay. An Overview of Language Models: Recent Developments and Outlook. *AP-SIPA Transactions on Signal and Information Processing*. 2024, roč. 13, č. 2. ISSN 2048-7703. Dostupné z DOI: 10.1561/116.00000010.
6. ZHOU, Ce; LI, Qian; LI, Chen; YU, Jun; LIU, Yixin; WANG, Guang-jing; ZHANG, Kai; JI, Cheng; YAN, Qiben; HE, Lifang; PENG, Hao; LI, Jianxin; WU, Jia; LIU, Ziwei; XIE, Pengtao; XIONG, Caiming; PEI, Jian; YU, Philip S.; SUN, Lichao. *A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT*. 2023. Dostupné z DOI: 10.48550/arXiv.2302.09419.
7. BROWN, Peter F.; DELLA PIETRA, Vincent J.; DESOUZA, Peter V.; LAI, Jenifer C.; MERCER, Robert L. Class-Based  $n$ -gram Models of Natural Language. *Computational Linguistics*. 1992, roč. 18, č. 4, s. 467–480. Dostupné také z: <https://aclanthology.org/J92-4003/>.

8. RADFORD, Alec; NARASIMHAN, Karthik; SALIMANS, Tim; SUTSKEVER, Ilya et al. Improving language understanding by generative pre-training. 2018. Dostupné také z: [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf).
9. DEVLIN, Jacob; CHANG, Ming-Wei; LEE, Kenton; TOUTANOVA, Kristina. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. Dostupné z arXiv: 1810.04805 [cs.CL].
10. HUGGINGFACE. *The AI community building the future*. [online]. 2025. [cit. 2025-04-01]. Dostupné z: <https://huggingface.co>.
11. WEBSTER, Jonathan J.; KIT, Chunyu. Tokenization as the Initial Phase in NLP. In: *COLING 1992 Volume 4: The 14th International Conference on Computational Linguistics*. 1992. Dostupné také z: <https://aclanthology.org/C92-4173/>.
12. MIELKE, Sabrina J.; ALYAFEAI, Zaid; SALESKY, Elizabeth; RAFFEL, Colin; DEY, Manan; GALLÉ, Matthias; RAJA, Arun; SI, Chenglei; LEE, Wilson Y.; SAGOT, Benoît; TAN, Samson. *Between words and characters: A Brief History of Open-Vocabulary Modeling and Tokenization in NLP*. 2021. Dostupné z arXiv: 2112.10508 [cs.CL].
13. JUN, Yennie. *All languages are NOT created (tokenized) equal* [online]. 2023. [cit. 2025-04-04]. Dostupné z: <https://web.archive.org/web/20230817165705/https://blog.yenniejun.com/p/all-languages-are-not-created-tokenized#expand>.
14. MIN, Bonan; ROSS, Hayley; SULEM, Elior; VEYSEH, Amir Pouran Ben; NGUYEN, Thien Huu; SAINZ, Oscar; AGIRRE, Eneko; HEINZ, Ilana; ROTH, Dan. *Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey*. 2021. Dostupné z arXiv: 2111.01243 [cs.CL].
15. ELAZAR, Yanai; KASSNER, Nora; RAVFOGEL, Shauli; RAVICHANDER, Abhilasha; HOVY, Eduard; SCHÜTZE, Hinrich; GOLDBERG, Yoav. Measuring and Improving Consistency in Pretrained Language Models. *Transactions of the Association for Computational Linguistics*. 2021, roč. 9, s. 1012–1031. ISSN 2307-387X. Dostupné z DOI: 10.1162/tacl\_a\_00410.
16. VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Lukasz; POLOSUKHIN, Illia. *Attention Is All You Need*. 2023. Dostupné z arXiv: 1706.03762 [cs.CL].

17. PFEIFFER, Jonas; KAMATH, Aishwarya; RÜCKLÉ, Andreas; CHO, Kyunghyun; GUREVYCH, Iryna. AdapterFusion: Non-Destructive Task Composition for Transfer Learning. In: MERLO, Paola; TIEDEMANN, Jorg; TSARFATY, Reut (ed.). *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Online: Association for Computational Linguistics, 2021, s. 487–503. Dostupné z DOI: 10.18653/v1/2021.eacl-main.39.
18. HOULSBY, Neil; GIURGIU, Andrei; JASTRZEBSKI, Stanislaw; MORRONE, Bruna; LARO USSILHE, Quentin de; GESMUNDO, Andrea; ATTARIYAN, Mona; GELLY, Sylvain. *Parameter-Efficient Transfer Learning for NLP*. 2019. Dostupné z arXiv: 1902.00751 [cs.LG].
19. BROWN, Tom B.; MANN, Benjamin; RYDER, Nick; SUBBIAH, Melanie; KAPLAN, Jared; DHARIWAL, Prafulla; NEELAKANTAN, Arvind; SHYAM, Pranav; SASTRY, Girish; ASKELL, Amanda; AGARWAL, Sandhini; HERBERT-VOSS, Ariel; KRUEGER, Gretchen; HENIGHAN, Tom; CHILD, Rewon; RAMESH, Aditya; ZIEGLER, Daniel M.; WU, Jeffrey; WINTER, Clemens; HESSE, Christopher; CHEN, Mark; SIGLER, Eric; LITWIN, Mateusz; GRAY, Scott; CHESS, Benjamin; CLARK, Jack; BERNER, Christopher; MCCANDLISH, Sam; RADFORD, Alec; SUTSKEVER, Ilya; AMODEI, Dario. *Language Models are Few-Shot Learners*. 2020. Dostupné z arXiv: 2005.14165 [cs.CL].
20. MINAEI, Shervin; MIKOLOV, Tomas; NIKZAD, Narjes; CHENAGHLU, Meysam; SOCHER, Richard; AMATRIAIN, Xavier; GAO, Jianfeng. *Large Language Models: A Survey*. 2025. Dostupné z arXiv: 2402.06196 [cs.CL].
21. NAVEED, Humza; KHAN, Asad Ullah; QIU, Shi; SAQIB, Muhammad; ANWAR, Saeed; USMAN, Muhammad; AKHTAR, Naveed; BARNES, Nick; MIAN, Ajmal. *A Comprehensive Overview of Large Language Models*. 2024. Dostupné z arXiv: 2307.06435 [cs.CL].
22. WARNER, Benjamin; CHAFFIN, Antoine; CLAVIÉ, Benjamin; WELLER, Orion; HALLSTRÖM, Oskar; TAGHADOUINI, Said; GALLAGHER, Alexis; BISWAS, Raja; LADHAK, Faisal; AARSEN, Tom; COOPER, Nathan; ADAMS, Griffin; HOWARD, Jeremy; WHITAKER, Jonathan; POLI, Iacopo. *Finally, a Replacement for BERT* [online]. 2024. [cit. 2025-04-04]. Dostupné z: <https://huggingface.co/blog/modernbert>.
23. LIU, Yinhan; OTT, Myle; GOYAL, Naman; DU, Jingfei; JOSHI, Mandar; CHEN, Danqi; LEVY, Omer; LEWIS, Mike; ZETTLEMOYER,

- Luke; STOYANOV, Veselin. *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. 2019. Dostupné z arXiv: 1907.11692 [cs.CL].
24. WARNER, Benjamin; CHAFFIN, Antoine; CLAVIÉ, Benjamin; WELLER, Orion; HALLSTRÖM, Oskar; TAGHADOUINI, Said; GALLAGHER, Alexis; BISWAS, Raja; LADHAK, Faisal; AARSEN, Tom; COOPER, Nathan; ADAMS, Griffin; HOWARD, Jeremy; POLI, Iacopo. *Smarter, Better, Faster, Longer: A Modern Bidirectional Encoder for Fast, Memory Efficient, and Long Context Finetuning and Inference*. 2024. Dostupné z arXiv: 2412.13663 [cs.CL].
25. NASUKAWA, Tetsuya; YI, Jeonghee. Sentiment analysis: capturing favorability using natural language processing. In: *Proceedings of the 2nd International Conference on Knowledge Capture*. Sanibel Island, FL, USA: Association for Computing Machinery, 2003, s. 70–77. K-CAP '03. ISBN 1581135831. Dostupné z DOI: 10.1145/945645.945658.
26. AQLAN, Ameen; BAIRAM, Dr. Manjula; NAIK, R Lakshman. A Study of Sentiment Analysis: Concepts, Techniques, and Challenges. In: 2019, s. 147–162. ISBN 978-1-4939-9044-3. Dostupné z DOI: 10.1007/978-981-13-6459-4\_16.
27. KUMAR, Sudhanshu; ROY, Partha Pratim; DOGRA, Debi Prosad; KIM, Byung-Gyu. *A Comprehensive Review on Sentiment Analysis: Tasks, Approaches and Applications*. 2023. Dostupné z arXiv: 2311.11250 [cs.AI].
28. MAO, Yanying; LIU, Qun; ZHANG, Yu. Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*. 2024, roč. 36, č. 4, s. 102048. ISSN 1319-1578. Dostupné z DOI: <https://doi.org/10.1016/j.jksuci.2024.102048>.
29. SURYAWANSHI, Nikhil. Sentiment analysis with machine learning and deep learning: A survey of techniques and applications. *International Journal of Science and Research Archive*. 2024, roč. 12, s. 005–015. Dostupné z DOI: 10.30574/ijjsra.2024.12.2.1205.
30. ZHANG, Lei; WANG, Shuai; LIU, Bing. *Deep Learning for Sentiment Analysis : A Survey*. 2018. Dostupné z arXiv: 1801.07883 [cs.CL].
31. ELMAN, Jeffrey L. Finding Structure in Time. *Cognitive Science*. 1990, roč. 14, č. 2, s. 179–211. Dostupné z DOI: [https://doi.org/10.1207/s15516709cog1402\\_1](https://doi.org/10.1207/s15516709cog1402_1).
32. HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long Short-Term Memory. *Neural Computation*. 1997, roč. 9, č. 8, s. 1735–1780. ISSN 0899-7667. Dostupné z DOI: 10.1162/neco.1997.9.8.1735.

33. BAHDANAU, Dzmitry; CHO, Kyunghyun; BENGIO, Yoshua. *Neural Machine Translation by Jointly Learning to Align and Translate*. 2016. Dostupné z arXiv: 1409.0473 [cs.CL].
34. WESTON, Jason; CHOPRA, Sumit; BORDES, Antoine. *Memory Networks*. 2015. Dostupné z arXiv: 1410.3916 [cs.AI].
35. SUKHBAATAR, Sainbayar; SZLAM, Arthur; WESTON, Jason; FERGUS, Rob. *End-To-End Memory Networks*. 2015. Dostupné z arXiv: 1503.08895 [cs.NE].
36. TAN, Xingwei; CAI, Yi; XU, Jingyun; LEUNG, Ho-Fung; CHEN, Wenhao; LI, Qing. Improving aspect-based sentiment analysis via aligning aspect embedding. *Neurocomputing*. 2020, roč. 383, s. 336–347. ISSN 0925-2312. Dostupné z DOI: <https://doi.org/10.1016/j.neucom.2019.12.035>.
37. XU, Qiannan; ZHU, Li; DAI, Tao; YAN, Chengbing. Aspect-based sentiment classification with multi-attention network. *Neurocomputing*. 2020, roč. 388, s. 135–143. ISSN 0925-2312. Dostupné z DOI: <https://doi.org/10.1016/j.neucom.2020.01.024>.
38. ZHANG, Wenxuan; LI, Xin; DENG, Yang; BING, Lidong; LAM, Wai. *A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges*. 2022. Dostupné z arXiv: 2203.01054 [cs.CL].
39. CHEBOLU, Siva Uday Sampreeth; DERNONCOURT, Franck; LIPKA, Nedim; SOLORIO, Thamar. OATS: A Challenge Dataset for Opinion Aspect Target Sentiment Joint Detection for Aspect-Based Sentiment Analysis. In: CALZOLARI, Nicoletta; KAN, Min-Yen; HOSTE, Veronique; LENCI, Alessandro; SAKTI, Sakriani; XUE, Nianwen (ed.). *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Torino, Italia: ELRA a ICCL, 2024, s. 12336–12347. Dostupné také z: <https://aclanthology.org/2024.lrec-main.1080/>.
40. ZHANG, Wenxuan; LI, Xin; DENG, Yang; BING, Lidong; LAM, Wai. Towards Generative Aspect-Based Sentiment Analysis. In: ZONG, Chengqing; XIA, Fei; LI, Wenjie; NAVIGLI, Roberto (ed.). *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Online: Association for Computational Linguistics, 2021, s. 504–510. Dostupné z DOI: 10.18653/v1/2021.acl-short.64.
41. LI, Yuncong; YIN, Cunxiang; ZHONG, Sheng-hua; PAN, Xu. *Multi-Instance Multi-Label Learning Networks for Aspect-Category Sentiment Analysis*. 2020. Dostupné z arXiv: 2010.02656 [cs.CL].

42. PENG, Haiyun; XU, Lu; BING, Lidong; HUANG, Fei; LU, Wei; SI, Luo. Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020, roč. 34, č. 05, s. 8600–8607. Dostupné z DOI: 10.1609/aaai.v34i05.6383.
43. WAN, Hai; YANG, Yufei; DU, Jianfeng; LIU, Yanan; QI, Kunxun; PAN, Jeff Z. Target-Aspect-Sentiment Joint Detection for Aspect-Based Sentiment Analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020, roč. 34, č. 05, s. 9122–9129. Dostupné z DOI: 10.1609/aaai.v34i05.6447.
44. KE, Cai; XIONG, Qingyu; WU, Chao; YI, Hualing; GAO, Min; CHEN, Jie. SimCPD: a simple framework for contrastive prompts of target-aspect-sentiment joint detection. *Neural Computing and Applications*. 2023, roč. 35, č. 22, s. 16577–16592. ISSN 1433-3058. Dostupné z DOI: 10.1007/s00521-023-08529-6.
45. CAI, Hongjie; XIA, Rui; YU, Jianfei. Aspect-Category-Opinion-Sentiment Quadruple Extraction with Implicit Aspects and Opinions. In: ZONG, Chengqing; XIA, Fei; LI, Wenjie; NAVIGLI, Roberto (ed.). *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online: Association for Computational Linguistics, 2021, s. 340–350. Dostupné z DOI: 10.18653/v1/2021.acl-long.29.
46. PONTIKI, Maria; GALANIS, Dimitris; PAVLOPOULOS, John; PAPAGEORGIOU, Harris; ANDROUTSOPOULOS, Ion; MANANDHAR, Suresh. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In: NAKOV, Preslav; ZESCH, Torsten (ed.). *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Dublin, Ireland: Association for Computational Linguistics, 2014, s. 27–35. Dostupné z DOI: 10.3115/v1/S14-2004.
47. JIANG, Qingnan; CHEN, Lei; XU, Ruifeng; AO, Xiang; YANG, Min. A Challenge Dataset and Effective Models for Aspect-Based Sentiment Analysis. In: INUI, Kentaro; JIANG, Jing; NG, Vincent; WAN, Xiao-jun (ed.). *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, 2019, s. 6280–6285. Dostupné z DOI: 10.18653/v1/D19-1654.
48. LIU, Bin; LIN, Tao; LI, Ming. Enhancing aspect-category sentiment analysis via syntactic data augmentation and knowledge enhancement. *Knowledge-Based Systems*. 2023, roč. 264, s. 110339. ISSN 0950-7051. Dostupné z DOI: <https://doi.org/10.1016/j.knosys.2023.110339>.



49. LI, Xin; BING, Lidong; ZHANG, Wenxuan; LAM, Wai. Exploiting BERT for End-to-End Aspect-based Sentiment Analysis. In: XU, Wei; RITTER, Alan; BALDWIN, Tim; RAHIMI, Afshin (ed.). *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*. Hong Kong, China: Association for Computational Linguistics, 2019, s. 34–41. Dostupné z DOI: 10.18653/v1/D19-5505.
50. TOLEDO-RONEN, Orith; ORBACH, Matan; KATZ, Yoav; SLONIM, Noam. Multi-Domain Targeted Sentiment Analysis. In: CARPUAT, Marine; MARNEFFE, Marie-Catherine de; MEZA RUIZ, Ivan Vladimir (ed.). *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Seattle, United States: Association for Computational Linguistics, 2022, s. 2751–2762. Dostupné z DOI: 10.18653/v1/2022.naacl-main.198.
51. ZHANG, Yice; YANG, Yifan; LI, Yihui; LIANG, Bin; CHEN, Shiwei; DANG, Yixue; YANG, Min; XU, Ruifeng. Boundary-Driven Table-Filling for Aspect Sentiment Triplet Extraction. In: GOLDBERG, Yoav; KOZAREVA, Zornitsa; ZHANG, Yue (ed.). *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022, s. 6485–6498. Dostupné z DOI: 10.18653/v1/2022.emnlp-main.435.
52. MUKHERJEE, Rajdeep; NAYAK, Tapas; BUTALA, Yash; BHATTACHARYA, Sourangshu; GOYAL, Pawan. PASTE: A Tagging-Free Decoding Framework Using Pointer Networks for Aspect Sentiment Triplet Extraction. In: MOENS, Marie-Francine; HUANG, Xuanjing; SPECIA, Lucia; YIH, Scott Wen-tau (ed.). *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Online a Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021, s. 9279–9291. Dostupné z DOI: 10.18653/v1/2021.emnlp-main.731.
53. WU, Zhengxuan; ONG, Desmond C. *Context-Guided BERT for Targeted Aspect-Based Sentiment Analysis*. 2020. Dostupné z arXiv: 2010.07523 [cs.CL].
54. SUN, Chi; HUANG, Luyao; QIU, Xipeng. Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. In: BURSTEIN, Jill; DORAN, Christy; SOLORIO, Thamar (ed.). *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, 2019, s. 380–385. Dostupné z DOI: 10.18653/v1/N19-1035.

55. LI, Jianwei; LI, Xianyong; DU, Yajun; FAN, Yongquan; CHEN, Xiaoliang; HUANG, Dong. An aspect-category-opinion-sentiment quadruple extraction with distance information for implicit sentiment analysis. *Inf. Technol. Contr.* 2023, roč. 52, č. 2, s. 445–456. Dostupné z DOI: 10.5755/j01.itc.52.2.32903.
56. SREESURYA, Ilayaraja; RATHI, Himani; JAIN, Pooja; JAIN, Tapan Kumar. Hypex: A Tool for Extracting Business Intelligence from Sentiment Analysis using Enhanced LSTM. *Multimedia Tools and Applications*. 2020, roč. 79, č. 47, s. 35641–35663. ISSN 1573-7721. Dostupné z DOI: 10.1007/s11042-020-08930-6.
57. KUMAR, Sudhanshu; YADAVA, Mahendra; ROY, Partha Pratim. Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. *Information Fusion*. 2019, roč. 52, s. 41–52. ISSN 1566-2535. Dostupné z DOI: <https://doi.org/10.1016/j.inffus.2018.11.001>.
58. BOSE, Rajesh; DEY, Raktim Kumar; ROY, Sandip; SARDDAR, Debabrata. Sentiment Analysis on Online Product Reviews. In: TUBA, Milan; AKASHE, Shyam; JOSHI, Amit (ed.). *Information and Communication Technology for Sustainable Development*. Singapore: Springer Singapore, 2020, s. 559–569. ISBN 978-981-13-7166-0. Dostupné z DOI: 10.1007/978-981-13-7166-0\_56.
59. WANG, Hongwei; AND, Lijuan Zheng. Sentiment classification of Chinese online reviews: a comparison of factors influencing performances. *Enterprise Information Systems*. 2016, roč. 10, č. 2, s. 228–244. Dostupné z DOI: 10.1080/17517575.2014.947635.
60. BAJ-ROGOWSKA, Anna. Sentiment analysis of Facebook posts: The Uber case. In: 2017, s. 391–395. Dostupné z DOI: 10.1109/INTELCIS.2017.8260068.
61. GUPTA, Rahul; SAMEER, Syed; MUPPAVARAPU, Harsha; ENDURI, Murali Krishna; ANAMALAMUDI, Satish. Sentiment Analysis on Zomato Reviews. In: *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)*. 2021, s. 34–38. Dostupné z DOI: 10.1109/CICN51697.2021.9574641.
62. BIRJALI, Marouane; KASRI, Mohammed; BENI-HSSANE, Abderrahim. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*. 2021, roč. 226, s. 107134. ISSN 0950-7051. Dostupné z DOI: <https://doi.org/10.1016/j.knosys.2021.107134>.
63. TADLE, Raul Cruz. FOMC minutes sentiments and their impact on financial markets. *Journal of Economics and Business*. 2022, roč. 118, č. C. Dostupné z DOI: 10.1016/j.jeconbus.2021.1.



64. DOH, Taeyoung; KIM, Sungil; YANG, Shu-Kuei. How You Say It Matters: Text Analysis of FOMC Statements Using Natural Language Processing. *The Federal Reserve Bank of Kansas City Economic Review*. 2021, roč. 106, č. 1. Dostupné z DOI: 10.18651/ER/v106n1DohKimYang.
65. XING, Frank Z.; CAMBRIA, Erik; WELSCH, Roy E. Natural language based financial forecasting: a survey. *Artificial Intelligence Review*. 2018, roč. 50, č. 1, s. 49–73. ISSN 1573-7462. Dostupné z DOI: 10.1007/s10462-017-9588-9.
66. ROGNONE, Lavinia; HYDE, Stuart; ZHANG, S. Sarah. News sentiment in the cryptocurrency market: An empirical comparison with Forex. *International Review of Financial Analysis*. 2020, roč. 69, s. 101462. ISSN 1057-5219. Dostupné z DOI: <https://doi.org/10.1016/j.irfa.2020.101462>.
67. KRAAIJEVELD, Olivier; DE SMEDT, Johannes. The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*. 2020, roč. 65, s. 101188. ISSN 1042-4431. Dostupné z DOI: <https://doi.org/10.1016/j.intfin.2020.101188>.
68. MANGURI, Kamaran; RAMADHAN, Rebaz; MOHAMMED AMIN, Pshko. Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks. *Kurdistan Journal of Applied Research*. 2020, roč. 5, s. 54–65. Dostupné z DOI: 10.24017/covid.8.
69. NASEEM, Usman; RAZZAK, Imran; KHUSHI, Matloob; EKLUND, Peter; KIM, Jinman. COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis. *IEEE Transactions on Computational Social Systems*. 2021, roč. PP, s. 1–13. Dostupné z DOI: 10.1109/TCSS.2021.3051189.
70. CHAKRABORTY, Koyel; BHATIA, Surbhi; BHATTACHARYYA, Siddhartha; PLATOS, Jan; BAG, Rajib; HASSANIEN, Aboul Ella. Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. *Applied Soft Computing*. 2020, roč. 97, s. 106754. ISSN 1568-4946. Dostupné z DOI: <https://doi.org/10.1016/j.asoc.2020.106754>.
71. ALAMOUDI, A.H.; ZAIDAN, B.B.; ZAIDAN, A.A.; ALBAHRI, O.S.; MOHAMMED, K.I.; MALIK, R.Q.; ALMAHDI, E.M.; CHYAD, M.A.; TAREQ, Z.; ALBAHRI, A.S.; HAMEED, Hamsa; ALAA, Musaab. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Systems with Applications*. 2021, roč. 167, s. 114155. ISSN 0957-4174. Dostupné z DOI: <https://doi.org/10.1016/j.eswa.2020.114155>.

72. ABD-ALRAZAQ, Alaa; ALHUWAIL, Dari; HOUSEH, Mowafa; HAMDI, Mounir; SHAH, Zubair. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *J Med Internet Res*. 2020, roč. 22, č. 4, e19016. ISSN 1438-8871. Dostupné z DOI: 10.2196/19016.
73. BAKER, Mohammed Rashad; TAHER, Yalmaz Najmaldin; JIHAD, Kamal h. Prediction of people sentiments on Twitter using machine learning classifiers during Russian-Ukrainian conflict. 2023. Dostupné z DOI: 10.21203/rs.3.rs-2410016/v1.
74. JIHAD, Kamal H.; BAKER, Mohammed Rashad; FARHAT, Mariem; FRIKHA, Mondher. Machine Learning-Based Social Media Text Analysis: Impact of the Rising Fuel Prices on Electric Vehicles. In: ABRAHAM, Ajith; HONG, Tzung-Pei; KOTTECHA, Ketan; MA, Kun; MANGHIRMALANI MISHRA, Pooja; GANDHI, Niketa (ed.). *Hybrid Intelligent Systems*. Cham: Springer Nature Switzerland, 2023, s. 625–635. ISBN 978-3-031-27409-1. Dostupné z DOI: 10.1007/978-3-031-27409-1\_57.
75. SHEN, Rong-Ping; ZHANG, Heng-Ru; YU, Hong; MIN, Fan. Sentiment based matrix factorization with reliability for recommendation. *Expert Systems with Applications*. 2019, roč. 135, s. 249–258. ISSN 0957-4174. Dostupné z DOI: <https://doi.org/10.1016/j.eswa.2019.06.001>.
76. LI, Hui; CUI, Jiangtao; SHEN, Bingqing; MA, Jianfeng. An intelligent movie recommendation system through group-level sentiment analysis in microblogs. *Neurocomputing*. 2016, roč. 210, s. 164–173. ISSN 0925-2312. Dostupné z DOI: <https://doi.org/10.1016/j.neucom.2015.09.134>. SI:Behavior Analysis In SN.
77. SERRANO-GUERRERO, Jesus; OLIVAS, Jose A.; ROMERO, Francisco P. A T1OWA and aspect-based model for customizing recommendations on eCommerce. *Applied Soft Computing*. 2020, roč. 97, s. 106768. ISSN 1568-4946. Dostupné z DOI: <https://doi.org/10.1016/j.asoc.2020.106768>.
78. RAY, Biswarup; GARAIN, Avishek; SARKAR, Ram. An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Applied Soft Computing*. 2021, roč. 98, s. 106935. ISSN 1568-4946. Dostupné z DOI: <https://doi.org/10.1016/j.asoc.2020.106935>.
79. FU, Xiangling; OUYANG, Tianxiong; YANG, Zaoli; LIU, Shaohui. A product ranking method combining the features–opinion pairs mining and interval-valued Pythagorean fuzzy sets. *Applied Soft Computing*. 2020, roč. 97, s. 106803. ISSN 1568-4946. Dostupné z DOI: <https://doi.org/10.1016/j.asoc.2020.106803>.

80. JIMÉNEZ-ZAFRA, Salud María; MARTÍN-VALDIVIA, M. Teresa; MOLINA-GONZÁLEZ, M. Dolores; UREÑA-LÓPEZ, L. Alfonso. How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain. *Artificial Intelligence in Medicine*. 2019, roč. 93, s. 50–57. ISSN 0933-3657. Dostupné z DOI: <https://doi.org/10.1016/j.artmed.2018.03.007>. Extracting and Processing of Rich Semantics from Medical Texts.
81. RAMÍREZ-TINOCO, Francisco Javier; ALOR-HERNÁNDEZ, Giner; SÁNCHEZ-CERVANTES, José Luis; SALAS-ZÁRATE, María del Pilar; VALENCIA-GARCÍA, Rafael. Use of Sentiment Analysis Techniques in Healthcare Domain. In: *Current Trends in Semantic Web Technologies: Theory and Practice*. Ed. ALOR-HERNÁNDEZ, Giner; SÁNCHEZ-CERVANTES, José Luis; RODRÍGUEZ-GONZÁLEZ, Alejandro; VALENCIA-GARCÍA, Rafael. Cham: Springer International Publishing, 2019, s. 189–212. ISBN 978-3-030-06149-4. Dostupné z DOI: 10.1007/978-3-030-06149-4\_8.
82. CLARK, Eric M.; JAMES, Ted; JONES, Chris A.; ALAPATI, Amulya; UKANDU, Promise; DANFORTH, Christopher M.; DODDS, Peter Sheridan. *A Sentiment Analysis of Breast Cancer Treatment Experiences and Healthcare Perceptions Across Twitter*. 2018. Dostupné z arXiv: 1805.09959 [cs.CL].
83. AYATA, Değer; YASLAN, Yusuf; KAMASAK, Mustafa E. Emotion Recognition from Multimodal Physiological Signals for Emotion Aware Healthcare Systems. *Journal of Medical and Biological Engineering*. 2020, roč. 40, č. 2, s. 149–157. ISSN 2199-4757. Dostupné z DOI: 10.1007/s40846-019-00505-7.
84. DESMET, Bart; HOSTE, Véronique. Emotion detection in suicide notes. *Expert Systems with Applications*. 2013, roč. 40, s. 6351–6358. Dostupné z DOI: 10.1016/j.eswa.2013.05.050.
85. WANG, Xinyu; ZHANG, Chunhong; JI, Yang; SUN, Li; WU, Leijia; BAO, Zhana. A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network. In: LI, Jiuyong; CAO, Longbing; WANG, Can; TAN, Kay Chen; LIU, Bo; PEI, Jian; TSENG, Vincent S. (ed.). *Trends and Applications in Knowledge Discovery and Data Mining*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, s. 201–213. ISBN 978-3-642-40319-4. Dostupné z DOI: 10.1007/978-3-642-40319-4\_18.
86. JUNG, Hyesil; PARK, Hyeoun-Ae; SONG, Tae-Min. Ontology-Based Approach to Social Data Sentiment Analysis: Detection of Adolescent Depression Signals. *J Med Internet Res*. 2017, roč. 19, č. 7, e259. ISSN 1438-8871. Dostupné z DOI: 10.2196/jmir.7452.

87. BAKER, Mohammed Rashad; MOHAMMED, Esraa Zeki; JIHAD, Kamal H. Prediction of Colon Cancer Related Tweets Using Deep Learning Models. In: ABRAHAM, Ajith; PLLANA, Sabri; CASALINO, Gabriella; MA, Kun; BAJAJ, Anu (ed.). *Intelligent Systems Design and Applications*. Cham: Springer Nature Switzerland, 2023, s. 522–532. ISBN 978-3-031-27440-4. Dostupné z DOI: 10.1007/978-3-031-27440-4\_50.
88. JIANG, Lin; SUZUKI, Yoshimi. Detecting hate speech from tweets for sentiment analysis. In: 2019, s. 671–676. Dostupné z DOI: 10.1109/ICSAI48974.2019.9010578.
89. BADJATIYA, Pinkesh; GUPTA, Shashank; GUPTA, Manish; VARMA, Vasudeva. Deep Learning for Hate Speech Detection in Tweets. In: *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*. ACM Press, 2017, s. 759–760. WWW '17 Companion. Dostupné z DOI: 10.1145/3041021.3054223.
90. RODRÍGUEZ, Axel; ARGUETA, Carlos; CHEN, Yi-Ling. Automatic Detection of Hate Speech on Facebook Using Sentiment and Emotion Analysis. In: *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*. 2019, s. 169–174. Dostupné z DOI: 10.1109/ICAIIIC.2019.8669073.
91. ZIDNY NAF'AN, Muhammad; BIMANTARA, Alhamda; LARASATI, Afiatari; RISONDANG, Ezar; NUGRAHA, Novanda. Sentiment Analysis of Cyberbullying on Instagram User Comments. *Journal of Data Science and Its Applications*. 2019, roč. 2, s. 88–98. Dostupné z DOI: 10.21108/jdsa.2019.2.20.
92. VESELOVSKÁ, Kateřina. *Sentiment analysis in Czech*. Sv. 16. Praha, Czechia: ÚFAL, Univerzita Karlova, 2017. Studies in Computational and Theoretical Linguistics. ISBN 978-80-88132-03-5. Dostupné také z: [https://ufal.mff.cuni.cz/books/preview/2017-veselovska\\_full.pdf](https://ufal.mff.cuni.cz/books/preview/2017-veselovska_full.pdf).
93. PONTIKI, Maria; GALANIS, Dimitris; PAPAGEORGIOU, Haris; ANDROUTSOPOULOS, Ion; MANANDHAR, Suresh; AL-SMADI, Mohammad; AL-AYYOUB, Mahmoud; ZHAO, Yanyan; QIN, Bing; DE CLERCQ, Orphée; HOSTE, Véronique; APIDIANAKI, Marianna; TANNIER, Xavier; LOUKACHEVITCH, Natalia; KOTELNIKOV, Evgeniy; BEL, Nuria; JIMÉNEZ-ZAFRA, Salud María; ERYİĞİT, Gülşen. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In: BERTHARD, Steven; CARPUAT, Marine; CER, Daniel; JURGENS, David; NAKOV, Preslav; ZESCH, Torsten (ed.). *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. San Diego, California: Association for Computational Linguistics, 2016, s. 19–30. Dostupné z DOI: 10.18653/v1/S16-1002.

94. ÇANO, Erion; BOJAR, Ondřej. Sentiment Analysis of Czech Texts: An Algorithmic Survey. In: *Proceedings of the 11th International Conference on Agents and Artificial Intelligence*. SCITEPRESS - Science a Technology Publications, 2019. Dostupné z DOI: 10.5220/0007695709730979.
95. PŘIBÁŇ, Pavel; PRAZAK, Ondřej. Improving Aspect-Based Sentiment with End-to-End Semantic Role Labeling Model. In: MITKOV, Ruslan; ANGELOVA, Galia (ed.). *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*. Varna, Bulgaria: INCOMA Ltd., Shoumen, Bulgaria, 2023, s. 888–897. Dostupné také z: <https://aclanthology.org/2023.ranlp-1.96/>.
96. CLARK, Kevin; LUONG, Minh-Thang; LE, Quoc V.; MANNING, Christopher D. *ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators*. 2020. Dostupné z arXiv: 2003.10555 [cs.CL].
97. HABERNAL, Ivan; PTÁČEK, Tomáš; STEINBERGER, Josef. Sentiment Analysis in Czech Social Media Using Supervised Machine Learning. In: BALAHUR, Alexandra; GOOT, Erik van der; MONTOYO, Andres (ed.). *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Atlanta, Georgia: Association for Computational Linguistics, 2013, s. 65–74. Dostupné také z: <https://aclanthology.org/W13-1609/>.
98. STEINBERGER, Josef; BRYCHCÍN, Tomáš; KONKOL, Michal. Aspect-Level Sentiment Analysis in Czech. In: BALAHUR, Alexandra; GOOT, Erik van der; STEINBERGER, Ralf; MONTOYO, Andres (ed.). *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Baltimore, Maryland: Association for Computational Linguistics, 2014, s. 24–30. Dostupné z DOI: 10.3115/v1/W14-2605.
99. ŠMÍD, Jakub; PŘIBÁŇ, Pavel; PRAZAK, Ondřej; KRÁL, Pavel. Czech Dataset for Complex Aspect-Based Sentiment Analysis Tasks. In: CALZOLARI, Nicoletta; KAN, Min-Yen; HOSTE, Veronique; LENCI, Alessandro; SAKTI, Sakriani; XUE, Nianwen (ed.). *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Torino, Italia: ELRA a ICCL, 2024, s. 4299–4310. Dostupné také z: <https://aclanthology.org/2024.lrec-main.384/>.
100. SIDO, Jakub; PRAŽÁK, Ondřej; PŘIBÁŇ, Pavel; PAŠEK, Jan; SEJÁK, Michal; KONOPÍK, Miloslav. *Czert – Czech BERT-like Model for Language Representation*. 2021. Dostupné z arXiv: 2103.13031 [cs.CL].

101. LAN, Zhenzhong; CHEN, Mingda; GOODMAN, Sebastian; GIMPEL, Kevin; SHARMA, Piyush; SORICUT, Radu. *ALBERT: A Lite BERT for Self-supervised Learning of Language Representations*. 2020. Dostupné z arXiv: 1909.11942 [cs.CL].
102. STRAKA, Milan; NÁPLAVA, Jakub; STRAKOVÁ, Jana; SAMUEL, David. RobeCzech: Czech RoBERTa, a Monolingual Contextualized Language Representation Model. In: *Text, Speech, and Dialogue*. Springer International Publishing, 2021, s. 197–209. ISBN 9783030835279. ISSN 1611-3349. Dostupné z DOI: 10.1007/978-3-030-83527-9\_17.
103. LEHEČKA, Jan; ŠVEC, Jan. Comparison of Czech Transformers on Text Classification Tasks. In: *Statistical Language and Speech Processing*. Springer International Publishing, 2021, s. 27–37. ISBN 9783030895792. ISSN 1611-3349. Dostupné z DOI: 10.1007/978-3-030-89579-2\_3.
104. AARSEN, Tom. *Dataset Card for "tomaarsen/setfit-absa-semeval-laptops"* [online]. 2023. [cit. 2025-04-15]. Dostupné z: <https://huggingface.co/datasets/tomaarsen/setfit-absa-semeval-laptops>.
105. AARSEN, Tom. *Dataset Card for "tomaarsen/setfit-absa-semeval-restaurants"* [online]. 2023. [cit. 2025-04-15]. Dostupné z: <https://huggingface.co/datasets/tomaarsen/setfit-absa-semeval-restaurants>.
106. STEINBERGER, Josef; BRYCHCÍN, Tomáš; KONKOL, Michal. *Restaurant Reviews CZ ABSA* [online]. 2014. [cit. 2025-04-16]. Dostupné z: <https://corpora.kiv.zcu.cz/sentiment/>. Přímý link: <https://corpora.kiv.zcu.cz/sentiment/CzechABSA-v2.zip>.
107. GOOGLE. *Gemini 2.0 Flash* [online]. 2025. [cit. 2025-05-02]. Dostupné z: <https://ai.google.dev/gemini-api/docs/models#gemini-2.0-flash>.
108. CHAUMOND, Julien; SANSEVIERO, Omar; LYSANDRE. *BERT base model (uncased)* [online]. 2024. [cit. 2025-04-25]. Dostupné z: <https://huggingface.co/google-bert/bert-base-uncased>.
109. CHAUMOND, Julien; SANSEVIERO, Omar; LYSANDRE. *BERT large model (uncased)* [online]. 2024. [cit. 2025-04-25]. Dostupné z: <https://huggingface.co/google-bert/bert-large-uncased>.
110. CHAUMOND, Julien; SANSEVIERO, Omar; LYSANDRE. *BERT base model (cased)* [online]. 2024. [cit. 2025-04-25]. Dostupné z: <https://huggingface.co/google-bert/bert-base-cased>.
111. CHAUMOND, Julien; SANSEVIERO, Omar; LYSANDRE. *BERT large model (cased)* [online]. 2024. [cit. 2025-04-25]. Dostupné z: <https://huggingface.co/google-bert/bert-large-cased>.



112. GOOGLE-RESEARCH. *Google Research github - BERT - bert/multilingual.md* [online]. 2024. [cit. 2025-04-26]. Dostupné z: <https://github.com/google-research/bert/blob/master/multilingual.md>.
113. CHAUMOND, Julien; SANSEVIERO, Omar; LYSANDRE. *BERT multilingual base model (cased)* [online]. 2024. [cit. 2025-04-26]. Dostupné z: <https://huggingface.co/google-bert/bert-base-multilingual-cased>.
114. CHAUMOND, Julien; LYSANDRE. *RoBERTa base model* [online]. 2024. [cit. 2025-04-26]. Dostupné z: <https://huggingface.co/FacebookAI/roberta-base>.
115. LAMPLE, Guillaume; CONNEAU, Alexis. *Cross-lingual Language Model Pretraining*. 2019. Dostupné z arXiv: 1901.07291 [cs.CL].
116. CONNEAU, Alexis; KHANDELWAL, Kartikay; GOYAL, Naman; CHAUDHARY, Vishrav; WENZKE, Guillaume; GUZMÁN, Francisco; GRAVE, Edouard; OTT, Myle; ZETTLEMOYER, Luke; STOYANOV, Veselin. *Unsupervised Cross-lingual Representation Learning at Scale*. 2020. Dostupné z arXiv: 1911.02116 [cs.CL].
117. CHAUMOND, Julien; LYSANDRE. *XLM-RoBERTa (base-sized model)* [online]. 2024. [cit. 2025-04-26]. Dostupné z: <https://huggingface.co/FacebookAI/xlm-roberta-base>.
118. CHAUMOND, Julien; LYSANDRE. *XLM-RoBERTa (large-sized model)* [online]. 2024. [cit. 2025-04-26]. Dostupné z: <https://huggingface.co/FacebookAI/xlm-roberta-large>.
119. HE, Pengcheng; LIU, Xiaodong; GAO, Jianfeng; CHEN, Weizhu. *DeBERTa: Decoding-enhanced BERT with Disentangled Attention*. 2021. Dostupné z arXiv: 2006.03654 [cs.CL].
120. HE, Pengcheng; GAO, Jianfeng; CHEN, Weizhu. *DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing*. 2023. Dostupné z arXiv: 2111.09543 [cs.CL].
121. MICROSOFT. *DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing* [online]. 2023. [cit. 2025-04-26]. Dostupné z: <https://huggingface.co/microsoft/deberta-v3-base>.
122. WARNER, Benjamin; CHAFFIN, Antoine; CLAVIÉ, Benjamin; WELLER, Orion; HALLSTRÖM, Oskar; TAGHADOUINI, Said; GALLAGHER, Alexis; BISWAS, Raja; LADHAK, Faisal; AARSEN, Tom; COOPER, Nathan; ADAMS, Griffin; HOWARD, Jeremy; POLI, Iacopo. *ModernBERT* [online]. 2024. [cit. 2025-04-26]. Dostupné z: <https://huggingface.co/answerdotai/ModernBERT-base>.

123. WARNER, Benjamin; CHAFFIN, Antoine; CLAVIÉ, Benjamin; WELLER, Orion; HALLSTRÖM, Oskar; TAGHADOUINI, Said; GALLAGHER, Alexis; BISWAS, Raja; LADHAK, Faisal; AARSEN, Tom; COOPER, Nathan; ADAMS, Griffin; HOWARD, Jeremy; POLI, Iacopo. *ModernBERT* [online]. 2024. [cit. 2025-04-26]. Dostupné z: <https://huggingface.co/answerdotai/ModernBERT-large>.
124. ARKHIPOV, Mikhail; TROFIMOVA, Maria; KURATOV, Yuri; SOROKIN, Alexey. Tuning Multilingual Transformers for Language-Specific Named Entity Recognition. In: ERJAVEC, Tomaž; MARCIŃCZUK, Michał; NAKOV, Preslav; PISKORSKI, Jakub; PIVOVAROVA, Lidia; ŠNAJDER, Jan; STEINBERGER, Josef; YANGARBER, Roman (ed.). *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*. Florence, Italy: Association for Computational Linguistics, 2019, s. 89–93. Dostupné z DOI: 10.18653/v1/W19-3712.
125. DEEPPAVLOV. *SlavicBERT* [online]. 2022. [cit. 2025-04-26]. Dostupné z: <https://huggingface.co/DeepPavlov/bert-base-bg-cs-pl-ru-cased>.
126. PRAŽÁK, Ondřej; PAŠEK, Jan; SIDO, Jakub. *CZERT-B* [online]. 2022. [cit. 2025-04-26]. Dostupné z: <https://huggingface.co/UWB-AIR/Czert-B-base-cased>.
127. STRAKA, Milan; NÁPLAVA, Jakub; STRAKOVÁ, Jana; SAMUEL, David. *Model Card for RobeCzech* [online]. 2024. [cit. 2025-05-01]. Dostupné z: <https://huggingface.co/ufal/robeczech-base>.
128. RAFFEL, Colin; SHAZEER, Noam; ROBERTS, Adam; LEE, Katherine; NARANG, Sharan; MATENA, Michael; ZHOU, Yanqi; LI, Wei; LIU, Peter J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*. 2020, roč. 21, č. 140, s. 1–67. Dostupné také z: <http://jmlr.org/papers/v21/20-074.html>.
129. LEHEČKA, Jan; ŠVEC, Jan. *FERNET-C5* [online]. 2024. [cit. 2025-05-01]. Dostupné z: <https://huggingface.co/fav-kky/FERNET-C5>.
130. LEHEČKA, Jan; ŠVEC, Jan. *FERNET-C5-RoBERTa* [online]. 2024. [cit. 2025-05-01]. Dostupné z: <https://huggingface.co/fav-kky/FERNET-C5-RoBERTa>.
131. GOOGLE. *Google Colaboratory* [online]. 2025. [cit. 2025-04-27]. Dostupné z: <https://colab.google>.
132. NVIDIA. *NVIDIA T4* [online]. 2020. [cit. 2025-04-27]. Dostupné z: <https://www.nvidia.com/en-us/data-center/tesla-t4/>.
133. HUGGINGFACE. *Datasets* [online]. 2025. [cit. 2025-04-27]. Dostupné z: <https://huggingface.co/docs/datasets/index>.



134. PANDAS. *pandas* [online]. 2024. [cit. 2025-04-27]. Dostupné z: <https://pandas.pydata.org>.
135. NUMPY. *NumPy* [online]. 2024. [cit. 2025-04-27]. Dostupné z: <https://numpy.org>.
136. HUGGINGFACE. *Transformers* [online]. 2025. [cit. 2025-04-27]. Dostupné z: <https://huggingface.co/docs/transformers/en/index>.
137. HUGGINGFACE. *Evaluate* [online]. 2025. [cit. 2025-04-27]. Dostupné z: <https://huggingface.co/docs/evaluate/index>.
138. PYTORCH. *PyTorch* [online]. 2025. [cit. 2025-04-27]. Dostupné z: <https://pytorch.org>.
139. HUGGINGFACE. *TrainingArguments* [online]. 2025. [cit. 2025-04-27]. Dostupné z: [https://huggingface.co/docs/transformers/v4.51.3/en/main\\_classes/trainer#transformers.TrainingArguments](https://huggingface.co/docs/transformers/v4.51.3/en/main_classes/trainer#transformers.TrainingArguments).
140. HUGGINGFACE. *Metric: accuracy* [online]. 2025. [cit. 2025-04-27]. Dostupné z: <https://huggingface.co/spaces/evaluate-metric/accuracy>.
141. HUGGINGFACE. *Metric: f1* [online]. 2025. [cit. 2025-04-27]. Dostupné z: <https://huggingface.co/spaces/evaluate-metric/f1>.
142. HOWARTH, Josh. *Number of Parameters in GPT-4 (Latest Data)* [online]. 2025. [cit. 2025-05-04]. Dostupné z: <https://explodingtopics.com/blog/gpt-parameters>.
143. OPENAI. *ChatGPT* [online]. 2025. [cit. 2025-05-04]. Dostupné z: <https://chatgpt.com>.
144. GOOGLE. *Google AI Studio* [online]. 2025. [cit. 2025-05-04]. Dostupné z: [https://aistudio.google.com/prompts/new\\_chat](https://aistudio.google.com/prompts/new_chat).
145. CLOUD, Alibaba. *Qwen* [online]. 2025. [cit. 2025-05-04]. Dostupné z: <https://chat.qwen.ai>.
146. DEEPSEEK. *Deepseek* [online]. 2025. [cit. 2025-05-04]. Dostupné z: <https://chat.deepseek.com>.
147. ANTHROPIC. *Claude* [online]. 2025. [cit. 2025-05-04]. Dostupné z: <https://claude.ai/chat>.
148. MICROSOFT. *Phi-4* [online]. 2025. [cit. 2025-05-04]. Dostupné z: <https://huggingface.co/microsoft/phi-4>.

## Obsah příloh

/	
└─ readme.txt.....	stručný popis obsahu média
└─ data.....	adresář s použitými daty
└─ src	
└─ models.....	zdrojové kódy natrénovaných modelů
└─ preproc.....	zdrojové kódy pro úpravu dat
└─ thesis.....	zdrojová forma práce ve formátu $\text{\LaTeX}$
└─ text .....	text práce
└─ thesis.pdf.....	text práce ve formátu PDF