



Kauno technologijos universitetas

Informatikos fakultetas

**LIETUVIŠKŲ ATSLIEPIMŲ INTERNETE SENTIMENTŲ
ANALIZĖ PASITELKIANČIAMS GILIUOSIUS KALBOS
MODELIUS**

Baigiamasis magistro studijų projektas

Brigita Vileikytė

Projekto autorė

Doc. Mantas Lukoševičius

Vadovas

Kaunas, 2024



Kauno technologijos universitetas

Informatikos fakultetas

**LIETUVIŠKŲ ATSILIEPIMŲ INTERNETE SENTIMENTŲ
ANALIZĖ PASITELKIANČIAUS GILIUOSIUS KALBOS
MODELIUS**

Baigiamasis magistro studijų projektas

Programų sistemų inžinerija (6211BX011)

Brigita Vileikytė

Projekto autorė

Doc. Mantas Lukoševičius

Vadovas

Dr. Jurgita Arnastauskaitė

Recenzentė

Kaunas, 2024



Kauno technologijos universitetas

Informatikos fakultetas

Brigita Vileikytė

Lietuviškų Atsiliepimų Internetė Sentimentų Analizė Pasitelkiant Giliuosius Kalbos Modelius

Akademinio sąžiningumo deklaracija

Patvirtinu, kad:

1. baigiamąjį projektą parengiau savarankiškai ir sąžiningai, nepažeisdama(s) kitų asmenų autoriaus ar kitų teisių, laikydamasi(s) Lietuvos Respublikos autorių teisių ir gretutinių teisių įstatymo nuostatų, Kauno technologijos universiteto (toliau – Universitetas) intelektualinės nuosavybės valdymo ir perdavimo nuostatų bei Universiteto akademinės etikos kodekse nustatytų etikos reikalavimų;
2. baigiamajame projekte visi pateikti duomenys ir tyrimų rezultatai yra teisingi ir gauti teisėtai, nei viena šio projekto dalis nėra plagijuota nuo jokių spausdintinių ar elektroninių šaltinių, visos baigiamojo projekto tekste pateiktos citatos ir nuorodos yra nurodytos literatūros sąrašė;
3. įstatymų nenumatytų piniginių sumų už baigiamąjį projektą ar jo dalis niekam nesu mokėjęs (-usi);
4. suprantu, kad išaiškėjus nesąžiningumo ar kitų asmenų teisių pažeidimo faktui, man bus taikomos akademinės nuobaudos pagal Universitete galiojančią tvarką ir būsiu pašalinta(s) iš Universiteto, o baigiamasis projektas gali būti pateiktas Akademinės etikos ir procedūrų kontrolieriaus tarnybai nagrinėjant galimą akademinės etikos pažeidimą.

Brigita Vileikytė

Patvirtinta elektroniniu būdu



Kauno technologijos universitetas

Informatikos fakultetas

Magistro projekto užduotis

Projekto tema

Lietuviškų Atsiliepimų Interneto Sentimentų Analizė Pasielkiant
Giliuosius Kalbos Modelius

Reikalavimai ir sąlygos
(tikslinti pavadinimą pagal
poreikį)

Vadovas / Vadovė

Doc. Mantas Lukoševičius

(vadovo pareigos, vardas, pavardė, parašas)

(data)

Brigita Vileikytė. Lietuviškų atsiliepimų internete sentimentų analizė pasitelkiant giliuosius kalbos modelius. Magistro baigiamasis projektas, vadovas doc. Mantas Lukoševičius; Kauno technologijos universitetas, Informatikos fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Programų sistemų inžinerija.

Raktiniai žodžiai: *Sentimentų analizė, Nuotaika, Lietuviškų atsiliepimų analizė, Prižiūrimasis mašininis mokymasis, Transformerio architektūros modeliai, BERT, T5, GPT-4, Daugiakalbiai didieji kalbos modeliai.*

Kaunas, 2024 m. 82 p.

Santrauka

Sentimentų analizė yra plačiai tiriama ir besivystanti uždavinys natūralios kalbos apdorojimo srityje. Naujaisi automatizuoti sprendimai šioje srityje sulaukia vis didesnio susidomėjimo iš įvairių įmonių bei žmonių internete. Nepaisant to, sentimentų analizė išlieka sudėtinga klasifikavimo užduotimi, ypač mažiau tirtose ir analizuotose kalbose. Norint objektyviai klasifikuoti sentimentus, būtina atsižvelgti į jų subjektyvumą ir mokėti atskirti tam tikrus bruožus, kaip kalbos netaisyklingumas, perkeltinės reikšmės ar ironija. Šiame darbe analizuojami lietuviški internetiniai atsiliepimai, apimantys įvairias sritis. Pateikiami sentimentų analizės rezultatai, kai šiai užduočiai naudojama palyginti maža duomenų imtis, ribotai ištirta kalba ir resursai. Atlikus esamų darbų ir klasikinių taikomų algoritmų analizę lietuvių natūralios kalbos apdorojimo srityje, pastebėta, kad klasikiniai mašininio mokymosi metodai turi ribotas klasifikavimo galimybes. Todėl šiame darbe detalčiau analizuojamas ir tiriamas daugiakalbių kalbos modelių pritaikymas lietuviškų sentimentų analizės uždavinyje.

Šiame darbe aptariamas sentimentų analizės uždavinys, jo poreikis ir panaudojimo galimybės. Aprašomi praktikoje ir moksliniuose tyrimuose taikomi metodai sentimentų analizei, jų panaudojimo galimybės bei apribojimai. Išsamiai analizuojami BERT ir T5 modeliai, kurie yra pritaikomi lietuvių kalbos dokumentų sentimentų analizei. Baigiamajame etape atliekama modelių rezultatų analizė ir palyginimas, pabrėžiant jų efektyvumą ir pritaikymo specifiką lietuvių kalbai.

Brigita Vileikytė. Sentiment Analysis of Lithuanian Online Reviews Using Deep Language Models. Master's Final Degree Project; supervisor Assoc. Prof. Dr. Mantas Lukoševičius; Faculty of Information Technology, Kaunas University of Technology.

Study field and area (study field group): Software Engineering.

Keywords: *Sentiment analysis, Emotion, Analysis of Lithuanian reviews, Supervised machine learning, Transformer models, BERT, T5, GPT-4, multilingual LLMs.*

Kaunas, 2024 m. 82 p.

Summary

The sentiment analysis task has been widely studied and researched in Natural Language Processing. Various companies and resellers on the internet are becoming more interested in the latest automated sentiment analysis solutions. However, this task remains challenging, particularly in languages like Lithuanian, which are complex and subjective regarding sentiment expression. In this paper, we delve into the challenges of sentiment analysis in the Lithuanian language, particularly in the context of a relatively small dataset and a language that has been limitedly studied. After a comprehensive review of existing research in Lithuanian NLP tasks, we found that classical machine learning approaches and classification algorithms have limited capabilities when analyzing and detecting sentiments in the Lithuanian language on limited resources. This paper describes and analyzes the performance of pre-trained multilingual Large Language Models in Lithuanian sentiment analysis.

This work explores sentiment analysis, its necessity, and its application possibilities. It thoroughly describes methods used in practice and research for sentiment analysis, including their possibilities and limitations. In this research, BERT and T5 models have been adapted for sentiment analysis of Lithuanian language documents. The paper concludes by analyzing and comparing the models' results, highlighting their effectiveness and practical implications for sentiment analysis in the Lithuanian language.

Turinys

Lentelių sąrašas	9
Paveikslų sąrašas	10
Terminų ir santrumpų sąrašas	11
Įvadas	13
1. Literatūros apžvalga	15
1.1. Sentimentų analizė	15
1.1.1. Sentimentų analizės apibrėžimas	15
1.1.2. Internetinių atsiliepimų sentimentų analizavimo aktualumas	17
1.2. Sentimento analizės užduotis	18
1.3. Sentimentų analizėje taikomi metodai	21
1.3.1. Sentimentų analizė: metodai paremti leksikologija	22
1.3.2. Sentimentų analizė: metodai paremti mašininio mokymusi	23
1.4. Sentimentų analizėje taikomų metodų apžvalga	29
1.4.1. Teksto žetonizacija	29
1.4.2. Teksto vektorizavimo metodai	30
1.5. Didieji kalbos modeliai taikomi sentimentų analizėje	33
1.5.1. BERT tipo modeliai	34
1.5.2. T5 tipo modeliai	37
1.6. Rezultatų vertinimo metrikos	41
2. Sistemos projektas	43
2.1. Sistemos veiklos sudėtis	43
2.1.1. Sistemos kontekstas	43
2.1.2. Sistemos veiklų padalijimas	44
2.1.3. Sistemos panaudojimo atvejai	44
2.2. Reikalavimai sistemos funkcionalumui	45
2.2.1. Funkciniai reikalavimai	45
2.2.2. Nefunkciniai reikalavimai	45
2.3. Sistemos vystymo procesas	46
2.4. Sistemos architektūros specifikacija	46
2.4.1. Esybių ryšių modelis	46
2.4.2. Statinis sistemos vaizdas	47
2.4.3. Sistemos išdėstymo vaizdas	50
2.5. Sistemos realizacija	51
3. Tyrimo metodai	54
3.1. Modelio kūrimo procesas	54
3.2. Duomenų imtis	55
3.2.1. Reikalavimai duomenims	55
3.2.2. Duomenų šaltiniai	56
3.2.3. Duomenų paruošimas ir analizė	56
3.3. distilBERT lietuviškų atsiliepimų sentimentų analizėje	59
3.3.1. Duomenų žetonizacija	60
3.3.2. Architektūros sluoksniai ir svarbiausi komponentai	60
3.3.3. Hiperparametrai ir rezultatai	61
4. Tyrimo rezultatai	64
4.1. distilBERT	64

4.2. ByT5	67
4.2.1. Modelis ir jo parametrai.....	67
4.2.2. Modelio rezultatai	68
4.3. GPT-4 Turbo	71
4.3.1. Taikomi apribojimai.....	71
4.3.2. Modelio rezultatai	71
4.4. Rezultatų palyginimas.....	73
Išvados.....	74
Literatūros sąrašas.....	76
Priedai	80
1 Priedas. Papildomos ByT5 modelio prognozių rezultatų diagramos.....	80
2 Priedas. Užklausų į OpenAI programinės sąsają programinis kodas	81

Lentelių sąrašas

1.4.1 lentelė. BoW metodų paremtas žodžių pasikartojimų išskyrimo pavyzdys	30
2.1.1 lentelė. Sistemos veiklos įvykių sąrašo detalizavimas	44
3.2.1 lentelė. Atsiliepimų įvertinimų priskyrimas sentimentų kategorijoms.....	55
3.2.2 lentelė. Duomenų rinkinių šaltinių informacija.....	56
3.2.3 lentelė. Duomenų pasiskirstymas apmokymo, validacijos bei testavimo duomenų imtyse.....	59
3.3.1 lentelė. Atsitiktinai parinktų testavimo įrašų distilBERT modelio prognozių pavyzdžiai.....	63
4.1.1 lentelė. distilBERT modelio prognozių maišaties matricos lentelė su pirmuoju duomenų rinkiniu.....	65
4.1.2 lentelė. distilBERT modelio prognozių maišaties matricos lentelė su antruoju duomenų rinkiniu	66
4.1.3 lentelė. distilBERT geriausio modelio 3 sentimentų kategorijų prognozių maišaties matricos lentelė.....	66
4.1.4 lentelė. distilBERT modelių našumo ir tikslumo įverčių apžvalga.....	66
4.1.5 lentelė. distilBERT modelių sentimentų analizės rezultatai	67
4.2.1 lentelė. ByT5 modelio prognozių maišaties matricos lentelė su pirmuoju duomenų rinkiniu ...	69
4.2.2 lentelė. ByT5 modelio prognozių maišaties matricos lentelė su antruoju duomenų rinkiniu ...	69
4.2.3 lentelė. ByT5 geriausio modelio 3 sentimentų kategorijų prognozių maišaties matricos lentelė	70
4.2.4 lentelė. ByT5 modelių našumo ir tikslumo įverčių apžvalga	70
4.2.5 lentelė. ByT5 modelių sentimentų analizės rezultatai	70
4.3.1 lentelė. GPT-4 modelio prognozių maišaties matricos lentelė	72
4.3.2 lentelė. GPT-4 modelio 3 sentimentų kategorijų prognozių maišaties matricos lentelė	72
4.3.3 lentelė. GPT-4 modelio sentimentų analizės rezultatai	72
4.4.1 lentelė. Geriausių sentimentų analizės modelių metrikų įverčiai	73

Paveikslų sąrašas

1.1.1 pav. Sąryšis tarp subjektyvumo, objektyvumo ir nuomonės sinonimų (sudaryta remiantis [5])	16
1.2.1 pav. Internetinių atsiliepimų įvertinimo pasiskirstymas (sudaryta remiantis [10])	20
1.3.1 pav. Giliuoju mokymusi paremtų mašininio modelių struktūros išskirtinumas (sudaryta remiantis [25])	26
1.3.2 pav. Sumažintas skaliarinės sandaugos dėmesio mechanizmas (sudaryta remiantis [28])	27
1.3.3 pav. Daugiagalvis dėmesio mechanizmas (sudaryta remiantis [28])	28
1.4.1 pav. CBOW ir Skip-gram modelių architektūros (sudaryta remiantis [35])	32
1.5.1 pav. BERT dvikryptė apmokymo modelio architektūra (sudaryta remiantis [39])	34
1.5.2 pav. Maskuoto kalbos modelio įvesties ir išvesties sekų atvaizdavimas (sudaryta remiantis [45])	36
1.5.3 pav. T5 modelio veikimo apibendrinimas (sudaryta remiantis [29])	38
1.5.4 pav. T5 modelio architektūra (sudaryta remiantis [49])	38
1.5.5 pav. Informacijos neprarandančios žetonizacijos algoritmo pavyzdys (sudaryta remiantis [41])	39
1.5.6 pav. T5 išankstinio mokymosi uždavinio pavyzdys (sudaryta remiantis [29])	40
2.1.1 pav. Prototipinės sentimentų analizės sistemos konteksto diagrama	43
2.1.2 pav. Prototipinės sentimentų analizės sistemos panaudojimo atvejų diagrama	44
2.4.1 pav. Sentimentų analizės sistemos esybių ryšių diagrama	47
2.4.2 pav. Bendra sistemos paketų diagrama	48
2.4.3 pav. Vartotojo sąsajos posistemės detalizuota paketų diagrama	49
2.4.4 pav. Serverinės sistemos dalies posistemės detalizuota <i>Domain</i> paketo diagrama	50
2.4.5 pav. Sistemos aplinkos diegimo diagrama	50
2.5.1 pav. Tinkamai suformatuoto .json dokumento tipo pavyzdys	51
2.5.2 pav. Tinkamai suformatuoto .csv dokumento tipo pavyzdys	51
2.5.3 pav. Sentimentų analizės sistemos vartotojo analizių sąrašo lango pavyzdys	52
2.5.4 pav. Atliktos sentimentų analizės pavyzdys	52
2.5.5 pav. Atliktos sentimentų analizės prognozių kategorijų pasiskirstymo pavyzdys	52
3.1.1 pav. Apibendrinta veiksmų seka sentimentų analizės modelio sukūrimui	54
3.2.1 pav. Duomenų įrašų pasiskirstymas pagal 5-ias sentimentų kategorijas	57
3.2.2 pav. Žodžių kiekio pasiskirstymo histograma duomenų imtyje	58
3.2.3 pav. Sentimentų kategorijų tolydinio skirstinio reikšmių diagrama	58
3.3.1 pav. Apibendrintas įvesties duomenų konvertavimui į modelio formatą procesas	60
3.3.2 pav. Pirmojo duomenų rinkinio distilBERT modelio mokymosi nuostolių įverčių kitimo diagrama	62
3.3.3 pav. distilBERT modelio, apmokyto su pirmuoju duomenų rinkiniu, individualių sentimentų kategorijų maišaties matricos	62
4.1.1 pav. Antrojo duomenų rinkinio distilBERT modelio mokymosi nuostolių įverčių kitimo diagrama	65
4.2.1 pav. Pirmojo duomenų rinkinio ByT5 modelio mokymosi nuostolių įverčių kitimo diagrama	68
34 pav. Antrojo duomenų rinkinio ByT5 modelio mokymosi nuostolių įverčių kitimo diagrama	80

Terminų ir santrumpų sąrašas

Santrumpos:

ANN (angl. *artificial neural network*) – dirbtinis neuroninis tinklas;

BoW (angl. *bag of words*) – žodžių rinkinio krepšelio metodas;

CRF (angl. *conditional random field*) – sąlyginiai atsitiktiniai laukai;

FNN (angl. *feedforward neural network*) – tiesioginis neuroninis tinklas;

ILA (angl. *inductive learning algorithm*) – induktyvaus mokymosi algoritmas;

LLM (angl. *large language model*) – didelės kalbos modelis;

LSA (angl. *latent semantic analysis*) – natūralios kalbos apdorojimo metodas suteikiantis galimybę analizuoti ryšius tarp įvairių tekstinių dokumento rinkinių ir su jais susijusių terminų;

LSTM (angl. *long short-term memory*) – neuroninis tinklas, skirtas spręsti ilgalaiškių priklausomybių problemas teksto sekuose;

MLM (angl. *masked language model*) – modelis, kuris mokosi prognozuoti atskirus žodžius, kai kai kurie žodžiai yra "užmaskuoti";

NB (angl. *Naive Bayes*) – Naiviojo Bajeso klasifikavimo algoritmas;

NLP (angl. *natural language processing*) – disciplina apimanti lingvistikos bei informatikos mokslus, tirianti įvairias pritaikymo galimybes natūralios kalbos sritims;

NSP (angl. *next sentence prediction*) – sekančio sakinio prognozavimo uždavinys, taikomas NLP uždavinių sprendimuose;

PCA (angl. *principal component analysis*) – pagrindinių komponentų analizės metodas;

POS (angl. *part of speech*) – kalbos dalys;

RNN (angl. *recurrent neural network*) – pasikartojantys neuroniniai tinklai;

SA (angl. *sentiment analysis*) – sentimentų analizė;

SVM (angl. *support vector machine*) – atraminių vektorių klasifikavimo metodas;

TF (angl. *term frequency*) – termino dažnis;

TF-IDF (angl. *term frequency-inverse document frequency*) – termino dažnis–atvirkštinis dokumento dažnis.

Terminai:

Sentimentas – emocinis kalbos ar požiūrio reiškimo atspalvis, naudojamas norint nustatyti, kaip subjektas jaučiasi ar vertina tam tikrus objektus, temas ar įvykius.

Sentimento poliariškumas – sentimento savybė turėti teigiamą, neutralią arba neigiamą emociją, naudojama sentimentų analizėje, siekiant nustatyti teksto ar kalbos turinio emocinį atspalvį.

Transformeris – modelis, kuris remiasi dėmesio mechanizmu, kuris leidžia sutelkti dėmesį į svarbiausias žodžių ar frazių dalis tekste, ir yra plačiai naudojamas ne tik natūralios kalbos apdorojime, bet ir vaizdų atpažinime ir kitose dirbtinio intelekto taikymo srityse.

Ivadas

Šiandieninis mūsų gyvenimas yra neįsivaizduojamas be ranka pasiekiamos, neribotos informacijos internete. Informacija tapo vienas svarbiausių objektų, kuriuo dalijamasi įvairiose gyvenimo srityse, bet neapsiribojant tik socialiniais tinklais, politinėmis naujienomis, žiniomis. Interneto vartotojams pasiekiamas informacijos turinys yra neatsiejamas nuo bendravimo su kitais žmonėmis bei savos nuomonės dalinimosi. Vis didėjantys nuomonių kupini duomenų resursai, kaip socialiniai tinklai, žinių straipsniuose paliekami komentarai, produktų bei paslaugų atsiliepimai, suteikia naujas galimybes pasitelkiant automatizuotus būdus gebėti suprasti kitų žmonių nuomones.

Internetinei erdvei perimant didelę vartotojų rinkos dalį, informacijos bei žinių dalijimasis tarp vartotojų tapo įprastos kasdienybės dalis. Šie duomenų srautai yra linkę įgyti vis didesnę reikšmę, kadangi žmonės vis dažniau pasikliauja kitų vartotojų nuomonėmis dėl įvairių sričių. Ši tendencija didina poreikį gebėti automatizuotu būdu analizuoti tekste perteikiamas emocijas. Įvairios įmonės bei organizacijos siekdamos išlaikyti teigiamą požiūrį į jų siūlomus produktus ar paslaugas yra suinteresuotos nuolat gauti teigiamus vartotojų įvertinimus. Tačiau žmonėms gali būti sunku apdoroti didelio kiekio tekstinę informaciją. Šiai problemai spręsti dažnai yra pasitelkiama sentimentų analizė.

Daugiau nei du dešimtmečius yra tobulinami bei atliekami įvairūs moksliniai tyrimai siekiant nustatyti pateikiamus sentimentus įvairaus tipo tekstuose. Apdorojant nestruktūrizuotus bei neapibrėžtus tekstinius duomenis susiduriama su įvairiomis problemomis: neapibrėžtos sąvokos; kalbos sintaksės bei leksikos subtilybės; vartotojų paliekami barbarizmai tekste; sarkazmas; informacijos konteksto trūkumas; taikomų metodų galimybių ribotumai bei kitos problemos. Be to, kaip minėta, teksto analizėje yra aptariamos subtilybės dažnai yra pritaikomos kiekvienai kalbai individualiai. Pastaruoju metu didžioji dalis tyrimų bei metodų yra atliekami remiantis anglų kalba. Sentimentų analizė lietuviškiems tekstams analizuoti nėra sulaukusi daug dėmesio. Tačiau automatizuoti būdai analizuoti lietuviškus vartotojų atsiliepimus galėtų būti pritaikomi įvairiuose įmonių veiklos analizių etapuose.

Šiame darbe analizuojamas ir tiriamas daugiakalbių kalbos modelių pritaikymas lietuvių kalbos atsiliepimų sentimentų analizei. Literatūros analizės skyriuje aprašomas sentimentų analizės uždavinys, jo poreikis ir panaudojimo galimybės. Taip pat pateikiama BERT ir T5 modelių apžvalga bei jų pritaikymo galimybės. Inžinerinėje darbo dalyje aprašoma prototipinė sistema, sukurta sentimentų analizės modelių išbandymui. Trečiajame skyriuje pateikiama išsami mokymo ir testavimo duomenų apžvalga bei aprašomas geriausias rezultatus pasiekusio sentimentų analizės modelio parengimo ir pritaikymo procesas. Ketvirtajame skyriuje išsamiai aprašomi distilBERT, ByT5 ir GPT-4 modelių sentimentų klasifikavimo rezultatai bei jų palyginimas.

Darbo metu atliktas tyrimas ir rezultatai buvo pristatyti 29-oje tarptautinėje IVUS konferencijoje 2024 metais.

Tikslas bei uždaviniai

Šio darbo tikslas – sukurti prototipinį teksto analizavimo modelį, kuris galėtų padėti aptikti perteikiamą sentimentą lietuviškuose tekstuose.

Šio darbo uždaviniai:

1. išanalizuoti taikomus sprendimus sentimentų analizės uždaviniuose;
2. surinkti bei suklasifikuoti duomenų imtis, reikalingas modeliams mokyti;

3. apdoroti duomenis;
4. sukurti prototipinį sentimentų analizės modelį;
5. sukurti prototipinę sistemą modelio demonstracijai;
6. aptarti bei palyginti sukurto modelio rezultatus.

1. Literatūros apžvalga

Šiame skyriuje išsamiai aprašomas sentimentų analizės uždavinys, aptariamas jo poreikis ir panaudojimo galimybės. Toliau pateikiama išsami BERT ir T5 modelių apžvalga, pabrėžiant jų struktūrinius ypatumus ir efektyvumą sentimentų analizės kontekste.

1.1. Sentimentų analizė

Šiuolaikinėje visuomenėje jau daugiau nei pusė pasaulio gyventojų turi prieigą prie interneto bei gali naudotis jo teikiamais privalumais. Žiniasklaida, reklamos, socialiniai tinklai bei internetinė prekyba tampa neatsiejama visuomenės kultūros dalis [1]. Augantis interneto naudotojų kiekis lemia ne tik didėjančią informacijos bei paslaugų pasiūlą interneto erdvėje, bet ir didėjančią informacijos kiekį suteikiamą pačių vartotojų. Dažnas žiniasklaidos puslapis leidžia vartotojams palikti savo nuomonę išreiškiančius komentarus po įvairiais publikuojamais naujienų straipsniais, socialiniuose tinkluose vartotojų komentarai bei jų peržiūros yra neatsiejamas sistemos funkcionalumas, o internetinės prekybos produktų atsiliepimai yra vienas iš lemiamų faktorių padedančių vartotojams priimti sprendimus dėl norimų pirkinių. Internetinė erdvė pasikeitė iš „skaitymo“ režimo į „skaitymo-redagavimo“ režimą [2]. Šis vis didėjantis vartotojų pateikiamų duomenų srautas tapo neatsiejama internetinio pasaulio dalimi.

Didelė dalis vartotojų vis labiau pasikliauja internete matomų atsiliepimų srautu. O po COVID-19 pandemijos vartotojų pasitikėjimas kitų vartotojų atsiliepimais išaugo daugiau nei 50% [3]. Įmonėms, internetinėje erdvėje, siūlančioms įvairius produktus ar paslaugas vis svarbiau sukurti ir išlaikyti teigiamą įvaizdį. Kadangi vartotojai vis dažniau ir drąsiau reiškia savo nuomonę, yra svarbu užtikrinti, kad klientų norai bei poreikiai būtų patenkinami. Problema atsiranda siekiant suvaldyti ir išanalizuoti tokius duomenų srautus įvairiuose šaltiniuose. Siekiant patenkinti klientų poreikius ypač svarbu suprasti vartotojų pateikiamas emocijas ar požiūrį dominančiu aspektu. Įvertinant nagrinėjamų duomenų kiekius, pasaulyje jau kurį laiką yra skiriama daug dėmesio tekstų analizavimo automatizavimui. Teksto sentimentų bei kategorijų analizė yra vienos pagrindinių tekstų analizės mokslo srities krypčių.

Šiame skyriuje pateikiamas sentimentų analizės apibrėžimas, , atsižvelgiant į šį darbą bei mokslinę literatūrą, kurioje nagrinėjami automatizuoti tekstų analizės metodai. Taip pat aptariamas sentimentų analizės aktualumas rinkoje. Be to, analizuojami atlikti tyrimai, mokslinės įžvalgos bei aptariami iškilę sunkumai ir problemos, susijusios su teksto sentimentų analizės tyrimais.

1.1.1. Sentimentų analizės apibrėžimas

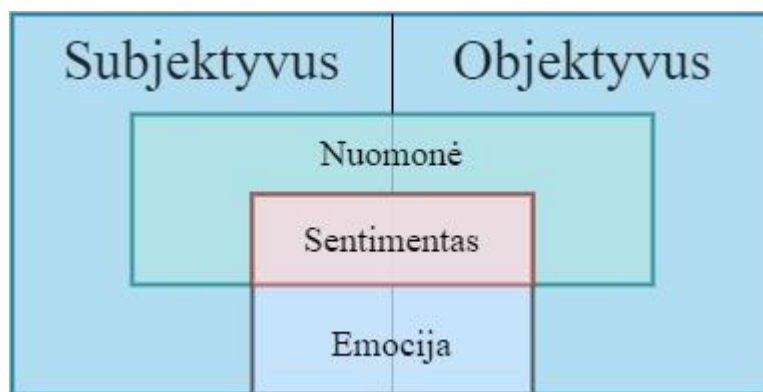
Žodis *sentimentas* lietuvių kalboje turi ne vieną apibrėžimą. Dažniausiai sutinkamas sentimento apibrėžimas yra pateikiamas sinonimais, kaip jausmas, požiūris ar nuotaika. Tačiau plečiantis tekstinės informacijos analizavimo sričiai, lietuvių kalbos naujažodžių duomenyne¹, terminui *sentimentas* taip pat yra suteikiamas emocinio kalbos atspalvio apibrėžimas [4]. Dažnai *sentimentas* yra pateikiamas kaip vienas iš žodžio ‘nuomonė’ sinonimų. Webster² anglų kalbos žodyne terminui *nuomonė* yra pateikiama daugiau nei 5 sinonimai. Nors žodžio *nuomonė* sinonimai savo prasme labai panašūs, vis dėlto analizuojant sentimentus yra naudinga gebėti juos atskirti [5]:

¹ Lietuvių kalbos naujažodžių duomenynas – internetinės svetainės adresas: <https://ekalba.lt/naujazodziai/naujienos>

² Webster anglų kalbos žodynas – internetinės svetainės adresas: <https://www.merriam-webster.com/>

- *nuomonė* – nusimanymas, nuovoka, supratimas. Ji gali būti suvokiama kaip tam tikras priimtinas nusistatymas, kuri dar gali būti diskutuotinas, formuojamas ar keičiamas;
- *požiūris* – tai konkreti pažiūra, nusistatymas. Jis gali būti apibrėžiamas, kaip subjektyvi nuomonė;
- *įsitikinimas* – sąmoningas tam tikros nuomonės priėmimas. Lietuvių kalbos visuotinėje enciklopedijoje³ apibrėžiamas kaip intelektualinis tikrumo būvis, kad asmens požiūris į tam tikrą objektą yra teisingas;
- *sentimentas* – nusistovėjusi nuomonė, atspindinti jausmus.

Visi šie terminai įprastai yra sietini su subjektyvumu. Tačiau analizuojant tekstus reikia gebėti įvertinti tiek subjektyviai, tiek objektyviai išreiškiamą būseną, kuri gali apimti nuomonę ar sentimentus. Svarbu suprasti, kad objektyvi nuomonė nebūtinai išreiškia sentimentą ar emociją. Taip pat subjektyvus pareiškimas, kuris perteikia nuomonę nebūtinai turi išreikšti emociją, todėl pareiškimas neturi ir sentimento (pvz.: „Aš tikiu dievu“). Galiausiai, tekste gali būti išreiškiamą emociją, tačiau nepateikiama jokia nuomonė ar sentimentas (pvz.: „Aš pagaliau atsipalaidavau!“) [5]. Apibendrinta šių terminų sąsaja yra atvaizduojama 1.1.1 pav. Tai reiškia, kad šie terminai gali būti vartotini, kai bandoma apibrėžti ne tik tai kas remiasi jausmais, emocijomis ar nuomone, bet ir būti naudojami siekiant perteikti objektyvų pareiškimą. Tai svarbu įvertinti siekiant suprasti, kodėl šie terminai yra laikomi sinonimais bei jų sąsaja teksto analizės tyrinėjimo mokslo srityje.



1.1.1 pav. Sąryšis tarp subjektyvumo, objektyvumo ir nuomonės sinonimų (sudaryta remiantis [5])

Sentimentų analizė (angl. *sentiment analysis*) – apibrėžiama kaip mokslo studijų sritis, kurioje analizuojami subjekto jausmai, nuostatos bei tam tikrų objektų vertinimas. *Nuomonių gavybos* (angl. *opinion mining*) terminą išskyrusios autorės B. Pang bei L. Lee *sentimentų analizę* laiko ekvivalentiu terminu vartotinu apibūdinti tą pačią mokslo sritį [6]. Šie terminai apima individo jausmų, nuomonės tyrbą į tam tikrą turinį, kaip žmones, prekes, paslaugas ar kita [7]. Šioje mokslo šakoje analizuojami įvairaus tipo tekstai, kurių apimtis gali kisti nuo keleto žodžių iki dokumentų pastraipų ar skyrių. Tekstų sentimentų analizė remiasi kalbotyros metodais, duomenų tyryba, statistika bei mašininiu mokymusi.

Pagrindinis sentimentų analizės tikslas yra gebėjimas nustatyti sentimento poliariškumą. Teksto tyrimų srityje terminas *poliariškumas* (kartais vadinamas valentingumu) naudojamas apibūdinti sentimento savybę, kuri gali būti teigiama, neutrali arba neigiama [6]. Analizuojant sentimentų poliariškumą svarbu suprasti sentimento tipą. Tipas yra neatsiejamas nuo sentimento vertinimo srities pasirinkimo, pavyzdžiui, vartotojų nuomonės tyrimo, lingvistikos ar psichologijos. Kadangi

³ Lietuvių kalbos visuotinė enciklopedija – internetinės svetainės adresas <https://www.vle.lt/>

sentimento identifikavimas gali būti sudėtingas net ir žmogui, o pati užduotis yra sudėtingesnė nei teksto kategorizavimas, papildoma informacija apie tekstą, temą ar paskirtį gali padėti geriau identifiukuoti teksto poliariškumą.

1.1.2. Internetinių atsiliepimų sentimentų analizavimo aktualumas

Nėra abejonių, kad interneto atsiradimas negrįžtamai pakeitė vartotojams aktualios bei dominančios informacijos gavimo procesą. Žmonės vis dažniau naujienas bei dominančią informaciją apie paslaugas, prekes, kitas gyvenimo sritis randa internete [1]. Oksfordo universiteto Reuters žiniasklaidos studijų instituto kasmetinio tyrimo duomenimis 2022-ais daugiau nei 65% amerikiečių naudojamas žinių šaltinis yra internetas (įskaitant ir socialinius tinklus). Tai yra labiausiai paplitęs informacijos pasiekiamumo būdas [1]. Tyrimo rezultatai pabrėžia, kad internetas yra neatsiejama kasdienio žmogaus gyvenimo dalis. Socialiniai tinklai, internetinės parduotuvės ir elektroniniai žinių portalai suteikia vartotojams galimybę interaktyviai naudotis sistemomis, skatina juos išreikšti savo nuomonę ir dalintis patirtimi. Tai generuoja didelį naujų duomenų kiekį, kuriame išreiškiamas įvairios vartotojų įžvalgos, nuotaikos ir nuomonės.

Minios išminties teorija (angl. *wisdom of the crowd*) teigia, kad grupė nepatyrusių asmenų yra pajėgi priimti teisingesnius sprendimus nei nedidelis kiekis srities ekspertų. Apibendrinta vartotojų nuomonė tampa svarbi tiek paslaugos tiekėjui, tiek gavėjui. Didėjantis informacijos kiekis suteikia galimybę sudaryti objektyvesnes įžvalgas dominančioje srityje bei priimti geresnius sprendimus [8]. Šios informacijos turiniu ypač suinteresuotos įmonės bei organizacijos reklamuojančios ar teikiančios paslaugas internetinėje erdvėje. Didelė dalis vartotojų vis labiau pasikliauja kitų žmonių pateikiamais atsiliepimais. Po *COVID-19* pandemijos vartotojų pasitikėjimas kitų naudotojų atsiliepimais išaugo daugiau nei 50% [3]. Daugiau nei 70% apklaustųjų 2022-aisiais metais atliktame įmonės *ReviewTrackers*⁴ tyrime teigia, kad blogi atsiliepimai iš kitų vartotojų pastūmėjo juos atsisakyti norimo ar juos dominančio produkto. Taip pat, pastebėta, kad vartotojai ypač vertina įmonės atsaką į jų paliktus atsiliepimus. Daugiau nei pusė tyrimo metu apklaustų asmenų teigia, kad galimai sugrįžtu pas buvusį paslaugos tiekėją, jei į jo pastabas būtų atsižvelgta ir atitinkamai sureaguota [3]. Šių apklausų rezultatai išryškina vartotojų nuomonės reikšmę internete. Tiek įmonėms, tiek paprastiesiems vartotojams priinama visuomenės atsiliepimų nuomonė suteikia didelę vertę priimant sprendimus. Vis dėlto, didelis neapibrėžtos struktūros vertinimų ir atsiliepimų kiekis apsunkina šios informacijos analizavimą ir struktūrizavimą.

Įmonės, siekiančios parduoti savo produktus visuomenei, skiria didelį dėmesį marketingui ir įvaizdžio tobulinimui. Vartotojo sprendimas pirkti prekę yra sudaromas iš daugelio skirtingų komponentų, tokių kaip asmens poreikis produktui, siūlomas kiekis, prekės klasė, pristatymo terminai ir užtikrinama kokybė, bei kitos dedamosios dalys. Kiekvienas nagrinėjamas aspektas leidžia vartotojui sumažinti alternatyvių produktų pasirinkimą. Produkto analizė atliekama ieškant papildomos informacijos internete ir priimant sprendimus remiantis kitų žmonių atsiliepimais. Galimybė lyginti paslaugas ir prekes tarp įvairių tiekėjų sparčiai didina konkurenciją tarp įmonių [9]. Jau įprastu patapęs atsiliepimų bei įvertinimų dalinimasis tarp vartotojų nulėmė, kad įmonės, siekdamos išlikti konkurencingos, privalo turėti kuo geresnį internetinį įvaizdį bei užtikrinti kiek įmanomą geresnę vartotojo patirtį. Atliktuose tyrimuose buvo pastebėta, kad kurios įmonės

⁴ Programinės įrangos, skirtos analizuoti klientų atsiliepimus, įmonė – internetinės svetainės adresas <https://www.reviewtrackers.com/>

suprasdamos paliekamų vartotojų atsiliepimų svarbą, pačios kuria netikrus atsiliepimus. Nustatyta, kad daugiau nei 16% paliekamų atsiliepimų svetainėje *Yelp*⁵ yra netikri atsiliepimai sukurti pačių verslų savininkų [10]. Vartotojų atsiliepimų valdymas bei pozityvų nuomonių formavimas tapo pagrindiniu marketingo srities darbo aspektu. Asmenys atsakingi už įmonės įvaizdžio formavimą analizuodami atsiliepimus daugiausiai dėmesio skiria vartotojų atsiliepimų kiekio užtikrinimui, jų pastovumui bei reitingui. Tačiau neretai pačių atsiliepimų turinys lieka neįvertinamas. Nors tyrimais nustatyta, kad vartotojai savo patirtis yra linkę aprašyti pateikiant daugiau įvairios informacijos, tačiau gali suteikti tik vieną apibrėžtą vertinimo įvertinimą. Pavyzdžiui, galbūt vienas klientas savo patirtį priskirs 1-os žvaigždutės įverčiui, kai tuo tarpu panašioje situacijoje atsidūręs kitas klientas šią situaciją galimai įvertins 3-mis žvaigždutėmis [10]. Todėl tiesioginis įverčių analizavimas yra labai neatsparus kritinėms reikšmės ir neturėtų būti naudojamas kaip vienintelis įmonės įvaizdžio analizavimo būdas.

Didelio kiekio atvirojo teksto atsiliepimų analizė yra sudėtinga ir monotoniška užduotis žmogui, tačiau ši informacija gali būti itin vertinga suinteresuotiems asmenims. Gebėjimas efektyviai analizuoti ir suprasti vartotojų poreikius gali lemiamai įtakoti įmonės sėkmę. Nuo 2000-ųjų pradėta tyrinėti kompiuterinės analizės galimybes laisvo teksto duomenims analizuoti. Dar 2002 metais buvo iškelta idėja, kad sentimentų nustatymas tekste yra sudėtingesnė užduotis nei kategorijos nustatymas, nes reikia gebėti įvertinti teksto turinį [11]. Pastebėta, kad tradiciniai, iki tol taikyti giliojo mokymosi kategorizavimo metodai nedavė patenkinamų rezultatų. Nuo to laiko susidomėjimas atviro teksto sentimentų klasifikavimo uždaviniu sparčiai išaugo, ir iki šių dienų stengiamasi atrasti įžvalgesnius bei tikslesnius sprendimus.

1.2. Sentimento analizės užduotis

Suprantant teksto sentimento nustatymo svarbą bei pritaikymo galimybes, aiškėja, kodėl natūralios kalbos apdorojimo (angl. *natural language processing*, NLP) bendruomenė skiria tiek daug dėmesio teksto sentimentų analizės tyrimams. Dėl didelio programų kiekio socialinėje, politinėje bei ekonominėje srityje, didžioji dalis sentimentų analizės tyrimų yra orientuoti siekiant nustatyti jei tam tikras *objektas* (žmogus, paslauga, produktas, renginys ir pan.) yra *laikomas*⁶ teigiamu ar neigiamu tam tikro *šaltinio* (žmogaus, įmonės, bendruomenės ir pan.). Sentimentų analizės tyrimai gali būti kategorizuojami pagal skirtingus kriterijus, kaip [5]:

- sentimento poliariškumo nustatymas;
- blogų ar gerų naujienų kategorizavimas;
- palaikančios ar nepalaikančios nuomonės identifikavimas;
- sutikimo ar prieštaravimo aptikimas;
- kitos alternatyvios kryptis.

Dėl plačios temų įvairovės bei tyrimo kryptių, sunku nustatyti esminius tyrinėjamus aspektus. Siekiant nustatyti autoriaus nuomonę bei tekste perteikiamą sentimentą, būtina identifikuoti pagrindinius analizuojamus aspektus. Mokslininkė B. Liu savo tyrimuose teigia, kad nuomonę gali apibrėžti remiantis 5-iomis dedamosiomis pateikiamomis (1) formulėje [12].

⁵ Verslo vertinimo ir apžvalgų svetainė – internetinės svetainės adresas: <https://www.yelp.ie/dublin>

⁶ *Laikomas* – terminas šiame kontekste naudojamas siekiant apibūdinti: įvertinimo procesą remiantis tam tikrais kriterijais, bei emocinę žmogaus, kuris pateikė vertinimą, būseną

$$(e_i, a_{ij}, s_{ijkl}, h_k, t_l); \quad (1)$$

čia:

- e_i – aptariamo objekto pavadinimas;
- a_{ij} – aptariamo e_i objekto aspektas;
- s_{ijkl} – perteikiamas sentimentas a_{ij} aspektui apie e_i objektą;
- h_k – nuomonės savininkas;
- t_l – laikas, kada nuomonė buvo išreikšta h_k subjekto.

Remiantis B. Liu teiginiais, kiekviena šios sąveikos dedamoji yra būtina, norint tinkamai įvertinti pateiktą nuomonę tekste. Laiko dedamoji yra svarbi norint gebėti analizuoti nuomonės kaitą, kaip keičiasi nuomonė apie aspektą keičiantis laikui. Taip pat svarbu gebėti identifikuoti ir nuomonės reiškėją. Pavyzdžiui sakinyje „*meras buvo mėgstamas miesto žmonių, tačiau sulaukdavo daug kritikos iš šalies valdžios*“ – svarbu gebėti nustatyti du nuomonės turėtojus, t.y. „*miesto žmonės*“ bei „*šalies valdžia*“. Sentimento dedamoji nuomonės apibrėžime gali įgyti teigiamos, neigiamos ar neutralios įverčius, tačiau taip pat gali būti analizuojama remiantis įvertinimais bei turėti intensyvumo lygius. Galiausiai, reikia įvertinti tai, kad šios priklausomybės negali apibrėžti visų galimų nuomonei priskiriamų aspektų. Šiuo išskaidymu nėra galimybės įvertinti perteikiamo konteksto esančiame tekste [12]. Pavyzdžiui sakinys „*šis butas yra per mažas didelei šeimai*“ ne teigia, kad aptariamas yra per mažas visiems asmenims. „*Didelė šeima*“ yra pagrindinis šio sakinio kontekstas, bet su aptartomis nuomonės dedamosiomis šis tekstas lieka neįvertinamas. Tačiau šio nuomonės apibrėžimo išplėtimas, įtraukiant dar daugiau priklausomybių, tik dar labiau apsunkina sprendžiamą problemą.

Šis nuomonės išskaidymas leidžia detaliau analizuoti sentimentą apibrėžimą. Kadangi sentimentas yra vertinamas kaip viena iš nuomonės dedamųjų – jis turi mažiau tarpusavio priklausomybių ir gali būti apibrėžiamas (2) formele [13].

$$(y, o, i); \quad (2)$$

čia:

- y – sentimentų tipas;
- o – sentimentų orientacija/poliariškumas;
- i – sentimentų intensyvumas.

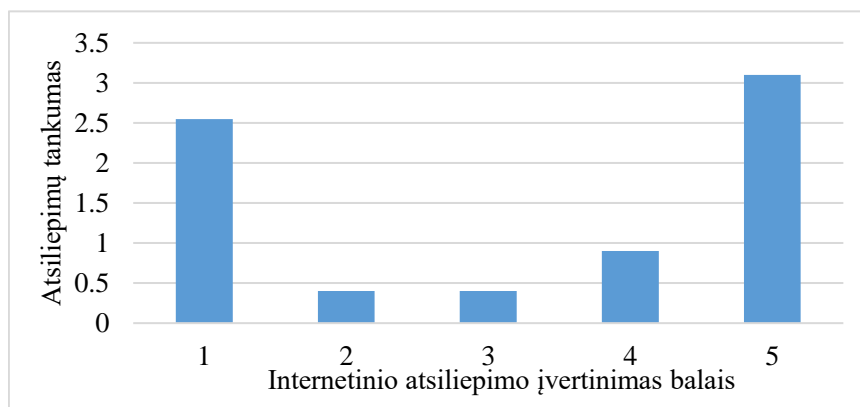
Siekiant nustatyti tekste perteikiamus sentimentus, būtina įvertinti šias tris pagrindines dedamąsias dalis, kurios apibūdina sentimentą. Sentimento tipas gali būti nustatomas remiantis analizės pagrindu, kuris gali apimti lingvistiką, psichologiją ar vartotojų patirties tyrimus. Daugiausiai duomenų ir dėmesio yra sulaukęs vartotojo patirties sentimentų analizavimas, kurį galima detaliau skirstyti į [13]:

- *racionalius sentimentus* – tai sentimentai neperteikiantys emocijos ir yra paremti tam tikrais įrodymais, tikėjimu ar kitais praktiškais aspektais;
- *emocionalius sentimentus* – tai neapčiuopiami ir emocijomis paremti išsireiškimai, kurių turinys ir prasmė yra labai priklausomas nuo autoriaus psichologinės būsenos.

Apibūdinant sentimentų orientaciją yra išskiriamos trys pagrindinės būsenos: teigiama, neigiama, neutrali. Neutralia sentimentų būsena galima nusakyti ir tai, kad analizuojamame tekste nėra perteikiamo emocinio sentimentų. Naujausiuose tyrimuose dažnai naudojami tiksliai apibrėžti įvertinimai siekiant išreikšti sentimentų intensyvumą bei orientaciją. Įprastai pritaikomas 5 lygių vertinimas (pvz.: 1-5 žvaigždės), kurį vėliau atliekant sentimentų analizę galima interpretuoti kaip [13]:

- emocijomis paremtą teigiamą sentimentą (+2 arba 5 žvaigždės);
- racionalumu paremtą teigiamą sentimentą (+1 arba 4 žvaigždės);
- neutralų ar emocijos neperteikiantį tekstą (0 arba 3 žvaigždės);
- racionalumu paremtą neigiamą sentimentą (-1 arba 2 žvaigždės);
- emocijomis paremtą neigiamą sentimentą (-2 arba 1 žvaigždė).

Toks atsiliėpimų įvertinimų suskirstymas palengvina reikiamų duomenų apdorojimą bei vertinimą, sumažina žmogiškųjų klaidų galimybes. Žinoma, kartais galima siekti identifikuoti detalesnius sentimentus, tačiau tai papildomai apsunkina nuomonių gavybos uždavinį, didėja paklaidų ir nuokrypių tikimybė. Suprantama, kad analizuoti pasitenkinimo taškus, paliekamą žvaigždučių įvertinimą, nykščio nuotaiką (angl. *thumbs-up/thumbs-down*) bei komentaro reitingus, yra kur kas paprasčiau nei analizuoti daugybę atviro teksto komentarų. Tačiau net ir analizuojant aiškiai kategorizuojamus įrašus yra pabrėžiama, kad vartotojų atsiliėpimai yra linkę nukrypti į kraštutinius. Pastebėta, kad vartotojai yra linkę palikti detalius įvertinimus, kai gauta prekė ar paslauga atitiko ar pranoko visus jų lūkesčius arba priešingai – paslauga ar prekė visiškai neatitiko kliento lūkesčių. Tai lemia, kad daugiausia turinio sulaukiama iš labiausiai pasitenkinusių bei nepasitenkinusių klientų (žr. 1.2.1 pav.) [10].



1.2.1 pav. Internetinių atsiliėpimų įvertinimo pasiskirstymas (sudaryta remiantis [10])

Dažnai analizė apsiriboja tik atsiliėpimų įverčių nagrinėjimu, tačiau tai lemia didelį vertingos informacijos praradimą [10]. Analizuojant ir vertinant atsiliėpimų turinį, galima geriau suprasti bendrą požiūrį į objektą. Žmonės, galintys išreikšti savo patirtį atviru tekstu, linkę pateikti daugiau įžvalgios informacijos, o bendras sentimentų pasiskirstymas paprastai yra artimesnis normaliajam pasiskirstymui.

Galiausiai, sentimentų intensyvumo dedamoji suteikia galimybę įvertinti tekstą pagal jo perduodamos emocijos stiprumą. Žmonės perteikdami savo jausmus bei emocijas tekste dažniausiai jas pateikia vienu iš dviejų būdų [13]:

- *remiantis jausmą perteikiančiais žodžiais*. Pavyzdžiui, sakinyje naudojamas žodis ‘geras’ turėtų mažesnę perteikiamą emocinę jausmą nei žodis ‘nuostabus’ ar ‘nepakartojamas’;
- *remiantis pagalbiniais stiprinančiais ar slopinančiais žodžiais*. Tokie išraiškos žodžiai pakeičia perteikiamo sentimentą kryptį ar intensyvumą. Pavyzdžiui žodžiai sustiprinantys sakinio intensyvumą yra ‘labai’, ‘nepaprastai’, ‘ypač’ ir kiti, o suprastinantys perteikiamą ryšį dažnai vartojami žodžiai, kaip ‘nežymiai’, ‘šiek tiek’, ‘truputį’, ‘iš dalies’ ir kiti.

Net identifikuojant svarbiausius elementus, sentimentų analizės uždavinys vis dar nėra trivialis užduotis. Per pastaruosius dešimtmečius ši NLP sritis plėtėsi įvairiais aspektais ir įgavo daug skirtingų formų. Dabar sentimentų analizė įprastai yra skirstoma pagal analizės detalumo lygį. Nors vieningo apibrėžimo sentimentų analizės turinio lygių išskaidymui nėra, literatūroje dažniausiai sutinkamas sentimentų analizės nagrinėjimas remiasi trimis lygiais [12]:

- *dokumento lygio* – analizuojama užduotis yra pateikiamo atsiliepimo apie produktą klasifikavimas. Sistema nustato ar pateikta apžvalga bendrai yra teigiama ar neigiama. Ši užduotis dažnai minima kaip dokumento lygio sentimentų klasifikavimas (angl. *document-level sentiment classification*) ir yra tinkama kai tekste yra apibūdinamas tik vienas objektas;
- *sakinio lygio* – analizuojama užduotis yra nustatyti, kokį sentimentą perteikia kiekvienas sakinyje. Sakinio lygio analizavime siekiama nustatyti vieną iš trijų emocijų: teigiamą, neigiamą ar neutralią;
- *objekto ar aspekto lygio* – siekiama identifikuoti subjektą bei perteikiamą apie jį emociją. Siekiama atrasti sentimentus apie tam tikrus objektus ir/ar jų aspektus. Pavyzdžiui sakinyje ‘Šio nešiojamo kompiuterio ekrano kokybė puiki, tačiau jo baterija yra silpna’ yra pateikiami du aspektai: *kompiuterio ekranas* bei *baterija*. Sentimentas apie ekrano kokybę yra teigiamas, o apie bateriją neigiamas. Šio lygio analizė suteikia galimybes struktūrizuoti analizuojamus tekstus išskiriant sentimentus bei aptariamus objektus. Tokia informacija gali būti naudojama tolimesnėje kokybinėje ir kiekybinėje analizėje.

Apibendrinant, sentimentų analizė yra plati teksto analizės sritis, kuri tiria tekstuose perteikiamą nuomonę bei orientaciją apie aptariamą objektą. Dažniausiai sentimentų analizės tikslas yra nustatyti perteikiamą emociją, kuri gali būti priskirta vienai iš trijų kategorijų: teigiama, neigiama arba neutrali. Tačiau egzistuoja ir kitokių skirstymo metodų. Teksto analizėje taikomi papildomi vertinimo kriterijai, siekiant nustatyti, apie kuriuos objektus kalbama, kokio intensyvumo yra emocija ar aptikti kitas emocijomis paremtas teksto charakteristikas. Atlikdami sentimentų analizę, svarbu įvertinti ne tik analizės lygmenį, bet ir iš anksto apsvarstyti analizuojamą tekstą: jo kalbos subtilybes, žodžių daugiareikšmiškumą, semantiką, emocijų perteikimo būdus bei kitus lingvistikos aspektus.

1.3. Sentimentų analizėje taikomi metodai

Sentimentų analizės klasifikavimo uždaviniai gali skirtis atsižvelgiant į analizuojamo teksto lygį, dominančią informaciją, taikymo sritį, sentimentą stiprumą ir kitus aspektus. Šie veiksniai lemia pasirinktus metodologinius sprendimus. Šiame skyriuje apžvelgiami plačiausiai paplitę ir žinomi metodai, pritaikomi sentimentų analizės uždaviniams spręsti.

Moksliniuose straipsniuose ir tyrimuose šiuolaikinėje sentimentų analizėje taikomi įvairūs metodų grupavimo būdai, tačiau paprastai yra išskiriami bent šie pagrindiniai metodai [14] [15]:

- leksika/žodynais pagrįsti metodai (angl. *lexicon-based approaches*);

- taisyklėmis grindžiami metodai (angl. *rule-based approaches*) [15];
- mašininio mokymosi pagrįsti metodai (angl. *machine-learning approaches*);
- hibridiniai metodai (angl. *hybrid approaches*).

Šie metodai gali būti analizuojami dar detaliau ir kiekvienas jų yra skirtas spręsti tam tikras problemas. Praktikoje dažnai išbandoma bent keletas sprendimo būdų iš pasirinktos metodų grupės. Taip pat dažnai aptinkamas hibridinis duomenų analizavimo modelis, kuris sujungia žodynų grįstus algoritmus su mašininio mokymosi galimybėmis. Tokie modeliai pasižymi aukštesniais tikslumo įvertinimais [16]. Šie hibridiniai metodai neturi apibrėžtos struktūros ir dažnai apima keletą skirtingų metodų kombinacijas, dažniausiai: žodynu grįstais, prižiūrimojo, neprižiūrimojo bei giliojo mašininio mokymosi metodais. Šiame skyriuje toliau detalizuojami metodai, kurie atskirai taikomi sentimentų analizės uždaviniui, ir prireikus gali būti kombinuojami kuriant hibridinį modelį.

1.3.1. Sentimentų analizė: metodai paremti leksikologija

Leksika ar (arba) raktiniais žodžiais pagrįsti metodai – tai metodai, kurie pritaiko vieną ar keletą žodyninių resursų tam, kad nustatytų teksto emociją. Šie metodai nustato teksto sentimentą poliškumą, remdamiesi žodžių semantine orientacija, kuri aptinkama analizuojamame tekste [15]. Paprastai raktinių žodžių paremti metodai grindžiami specialiai sudarytais žodynais, dažnai sukurti remiantis įvairių terminų sinonimais ir antonimais, o metodai remiasi žodžiais, kurie priklauso teigiamų ar neigiamų kategorijų grupėms [17]. Šie žodynai nuolat papildomi ir pritaikomi įvairioms analizuojamų temų kryptims. Naudojami žodynai yra įvairių tipų, daugiausiai žodynų yra pritaikyti anglų kalbai. Leksikoniniai rinkiniai gali turėti rankomis įvertintus žodžių junginius tarp -5 ir +5 balų, taip pat kai kurie žodynai yra sugeneruojami automatizuotai, kiti remiasi žmonių komentarais ir socialiniuose tinkluose naudojamais grotelėmis (angl. *hashtags*), arba sudaromi remiantis kalbos žodynu ir įvertinant įvairias žodžių formas, jas klasifikuojant pagal sinonimus [16]. Tokio tipo žodynus naudojantys analizės metodai mokslinėje literatūroje yra kategorizuojami pagal jų taikomas metodologijas siekiant nustatyti tekste perteikiamą sentimentą.

1.3.1.1. Raktažodžiais grindžiami metodai

Raktažodžiais grindžiami metodai (angl. *keyword-based approaches*) yra paremti iš anksto apibrėžiamais terminais, kuriuos aptikus yra nustatoma emocijos kategorija. Šio metodo pagrindas remiasi paprastu algoritmu, kuris bando aptikti aprašytus aktyvius terminus tekste bei antraštėse [14]. Remiantis žodžių atitikmenų skaičiumi, apskaičiuojamas pasikartojimų dažnumas, o jiems priskiriami pozityvūs arba negatyvūs balai, ir taip pateikiamas galutinis nustatytas sentimentas.

1.3.1.2. Ontologija grindžiami metodai

Ontologija grindžiami metodai (angl. *ontology-based approaches*) yra paremti žiniomis grįstais (angl. *knowledge based*) metodais. Ontologija apibūdina tyrinėjamos krypties žinias kaip tam tikrų aspektų hierarchiją [18]. Siekiant pritaikyti šiuos metodus, sudaromi apibrėžto formato žodynai, kurie padeda dalintis bendromis žiniomis tarp skirtingų sričių ekspertų. Ontologiniai modeliai gali būti saugomi OWL/XML arba RDF/XML formatuose. Kadangi šie modeliai yra paremti apibrėžtos hierarchijos struktūros sukūrimu – jie suteikia galimybę sujungti keletą modelių į vieną su sąlyga, kad jie aprašo tą pačią nagrinėjamą sritį [18]. *EmotiNet* modelis yra vienas iš ontologija grįstų metodų pavyzdžių, paremtas bendru supratimu (angl. *commonsense knowledge*) apie gyvenimiškas situacijas. Šiame modelyje kiekviena situacija yra sukuriama kaip tam tikrų veiksmų ar įvykių seka, kuriems priskiriami atitinkami emocijų žymenys įvairiose situacijose. Sudarytos įvykių grandinės yra

veiksmų sekos, kurios sužadina tam tikros emocijos aptikimą pasirinktam subjektui [19]. Šie metodai išsiskiria galimybe nustatyti emocijas iš platesnio aptariamo intervalo ir gali aptikti įvairias emocijas, tokias kaip pyktis, pasimetimas, baimė, gėda, laimė, nuostaba bei kitos.

1.3.1.3. Statistiniai metodai

Statistiniai metodai (angl. *statistical approach*) įprastai yra paremti latentine semantine analize (angl. *latent semantic analysis, LSA*), kurios metu sukuriama sąvokų žodyna, apibūdinantys įvairių sričių tekstinius rinkinius ir su jais susijusius terminus. Šie metodai yra grindžiami ryšių tarp tam tikrų dokumentų ir juose minimų terminų analize, siekiant atrasti prasmingus pasikartojančius šablonus [14]. Tradiciniai statistiniai metodai laikomi silpnais semantinę prasmę analizuojančiais metodais. Įprasti algoritmai gali veikti patenkinamai tik su didele duomenų imtimi ir reliatyviai dideliais duomenų kiekiais, nes jie nėra pajėgūs tikslingai identifikuoti vartotojų tekstų, kurie apima tik frazes ar pavienius sakinius. Todėl dažnai statistiniai metodai yra apjungiami su kitais sentimentų analizėje taikomais metodais, taip suteikiant galimybę patobulintam metodui aptikti ne tik aktyvinančius žodžius ir juos priskirti aprašytai emocijai [20]. Apjungti statistiniai modeliai išsiskiria gebėjimu į savo analizės įtraukti įvairius elementus, tokius kaip savavališkai parenkami, tarpusavyje susiję žodžiai, skyrybos ženklai, taip pat ir pasikartojančių žodžių dažnius. Toks metodų pritaikymas leidžia gauti išsamesnius ir tikslesnius sprendimus.

1.3.1.4. Leksika paremtų metodų apibendrinimas

Žodyniniai metodai plačiai taikomi įvairiuose NLP uždaviniuose. Tačiau bandant nustatyti teksto sentimentą naudojant leksikoninius metodus dažnai susiduriama su tam tikrais žodyninių metodų trūkumais bei apribojimais [16]:

- aptinkami dviprasmiškumai raktinių žodžių aprašymuose su kuriais yra siejamos emocijos. Išskyrus žodžius, kurie savo prasme yra priskiriami emocijoms, dauguma žodžių gali pakeisti savo prasmę, kai sakiniai yra paremti ironija ar ciniškumu;
- nesugebama aptikti emocijų sakiniuose, kuriuose neaptinkami raktiniai žodžiai. Taikant žodyninius metodus, kai sakiniuose nerandamas joks raktažodis – sakiniai tampa klasifikuojami, kaip neturintys jokio sentimentų;
- neatsižvelgiama į lingvistinę informaciją, tokią kaip sakinių sintaksė, struktūra, semantika, subjektas, teksto turinys ir kiti kalbos aspektai;
- kuriami sentimentų žodynai įprastai pritaikyti tik tam tikroms temų kategorijoms. Tai apsunkina nuomonių gavybą, kadangi remiantis tais pačiais žodynais skirtingose tematikose aprašyti žodžiai gali turėti skirtingus sentimentus [17].

Apibendrinant, žodyniniai metodai yra daugiausiai paremti statine teksto analize ir remiasi raktinių žodžių aptikimu pateikiamuose dokumentuose. Leksika grįsti metodai turi nemažai trūkumų ir juos taikant praktikoje jie dažnai yra modifikuojami arba apjungiami su kitais sentimentų analizės sprendžiančiais metodais. Šie pakeitimai būtini norint pagerinti sentimentų analizės rezultatų tikslumą ir pritaikyti metodus prie sudėtingesnių kalbos ypatybių atpažinimo.

1.3.2. Sentimentų analizė: metodai paremti mašininio mokymusi

Mašininis mokymasis – tai mokslo sritis, kurioje ypatingas dėmesys yra skiriamas induktyviems algoritmams (angl. *induction logic algorithms, ILA*) bei kitiems metodams, kurie turi galimybę „mokyti“ [21]. Šioje mokslo kryptyje tiriamos įvairios algoritmų pritaikymo galimybės, kai jie gali

mokytis iš pateikiamų duomenų [21]. Tokio tipo algoritmai veikia kurdami modelius, kurie, priklausomai nuo jiems paduodamų reikšmių, gali prognozuoti arba priimti sprendimus apie pateiktus duomenis, o ne tiesiogiai remtis nurodytomis instrukcijomis [14]. Orientuojantis į sentimentų analizės uždavinį, mašininio mokymosi algoritmai yra pritaikomi aptikti emocijas tekste. Šie metodai gali būti skirstomi į:

- prižiūrimojo tipo (angl. *supervised machine learning*);
- neprižiūrimojo tipo (angl. *unsupervised machine learning*).

1.3.2.1. Prižiūrimojo mašininio mokymosi metodai

Prižiūrimojo mašininio mokymosi metodai – yra paremti duomenimis, kurie yra iš anksto suklasifikuoti ar kitaip pažymėti prieš pradėdant mokymosi procesą [22]. Šie metodai mokymosi metu naudoja sužymėtų duomenų rinkinį ir geba atlikti prognozes duomenims, kurių neturėjo apmokymo metu. Analizuodami mokymosi duomenis, prižiūrimojo tipo mašininio mokymosi metodai geba sukurti atitinkamą funkciją, kuri vėliau yra taikoma naujų, anksčiau nematytų duomenų apdorojimui [22]. Šie metodai dažniausiai pritaikomi klasifikacijos, regresijos bei reitingavimo uždaviniuose.

Kadangi sentimentų analizė yra laikoma teksto klasifikavimo uždaviniu, galima teigti, kad prižiūrimojo mašininio mokymosi metodai yra tinkami nustatyti tekste perteikiamą sentimentą. Šie metodai naudoja iš anksto sužymėtus duomenis, kad išmokytų atpažinti ir klasifikuoti tekstą pagal jo sentimentą. Pirmieji ir labiausiai paplitę metodai sentimentų analizės uždaviniui yra Naivaus Bajeso (angl. *Naive Bayes*, NB) bei atramos vektorių (angl. *Support Vector Machine*, SVM) metodai [12]. Šie metodai pirmą kartą buvo pritaikyti sentimentų analizės uždaviniui spręsti dar 2002. B. Pang, L. Lee bei S. Vaithyanathan'as teikė publikaciją, kurioje aptariama, kaip naudojant (angl. *unigrams*) kaip požymius teksto analizėje, šie gali būti pritaikomi NB ir SVM [11]. Po šios publikacijos NLP srityje, sprendžiant sentimentų analizės uždavinį, buvo pritaikyta daugybė įvairių klasifikavimo metodų. Kiti žymūs darbai, taikantys prižiūrimojo mašininio mokymosi algoritmus, naudoja tiek kategorinius, tiek dimensinius metodus, formuojant emocijų aptikimo modelius. Šie modeliai leidžia atpažinti daugiau nei tris įprastus sentimentus; pavyzdžiui, kai kurie modeliai gali aptikti meilės emociją arba aštuonias dvipoles emocijas, remiantis R. Plutchick'o emocijų teorija. Tokiems metodams galima pritaikyti tolimojo stebėjimo (angl. *distant supervision*) metodus [14]. Siekiant pritaikyti šiuos algoritmus nuomonių svarbu išanalizuoti tekstą pagal pasirinktus požymius. Sentimentų klasifikavime galima išskirti ne vieną požymį, kuris padėtų aptikti teksto sentimentą. Keletas pavyzdžių [12]:

- *terminai ir jų pasikartojimo dažnis* – šie požymiai yra paremti unigramų ir jų sekų, vadinamų n-gramomis (angl. *n-grams*), pasikartojimų kiekiu tekste;
- *kalbos dalys* (angl. *part of speech*, POS) – šie požymiai remiasi lingvistikos kalbos dalių atpažinimu. Pavyzdžiui, pabrėžtina, kad būdvardžiai įprastai yra svarbesni bandant atpažinti esamą emociją tekste;
- *sentimentų žodžiai bei frazės* – tai tekste esantys žodžiai, kurie savo reikšme perteikia emociją (pvz.: *nuostabus, prastas, nekęsti, mylėti*);
- *emocijos reikšmės pakeitimas* (angl. *sentiment shifter*) – tai išsireiškimai, frazės, kurie pakeičia numanomo sentimento reikšmę tekste. Tai gali būti įvairūs priešdėliai (pvz.: *nenoriu, nepatinka*), neiginiai bei kitos kalbos dalys.

Klasikiniai prižiūravimo tipo mašininio mokymosi metodai yra gana plačiai naudojami bei pritaikomi sentimentų analizės aptikimui tekste. Tačiau jų taikyme dažnai susiduriama su problemomis, kaip:

- *emocijų rodiklių apibrėžimas* (angl. *emotion indicator determination*) – nors šie metodai gali automatizuotai atrasti tikimybes tarp įvairių požymių ir su jais susietų emocijų, vis dėlto požymių atradimas yra paremtas leksikologija ir tekstų apdorojimu. Dėl šių priežasčių taikant klasikinius klasifikavimo algoritmus, dažnai susiduriama su tomis pačiomis problemomis, kaip ir žodyniniuose metoduose [16];
- *supaprastintos emocijų kategorijos* (angl. *over-simplified emotion categories*) – dauguma modelių yra pritaikomi ir gerai geba atpažinti tik dalį, o neretai vos tris sentimentus (teigiamą, neigiamą bei neutralų), todėl modeliai gali būti sunkiai pritaikomi praktikoje [23];
- *suklasifikuotų duomenų imtys* – prižiūravimo mokymosi metodai yra paremti didelėmis bei gerai suklasifikuotų mokymosi duomenų imtimis, o analizuojant didelius kiekius teksto neretai susiduriama su duomenų klasifikavimo problema. Neretai pasirenkamas gretutinis rodiklis iš įvairių tekstų tipų, pagal kurį priskiriama kategorija, kuri neatitinka perteikiamo sentimentu [14]. Šis darbas taip pat yra sunkus atlikti ir žmonėms dėl didelio apdorojimo reikalingo duomenų kiekio ir sentimentų neapibrėžtumo.

1.3.2.2. Neprižiūravimo mašininio mokymosi metodai

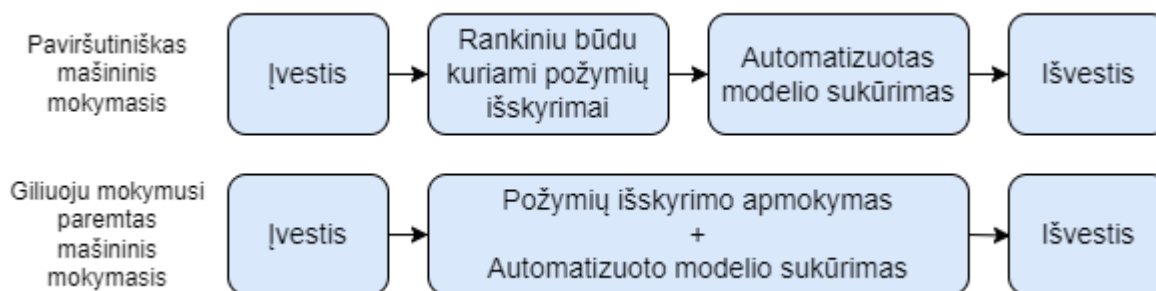
Neprižiūravimo mašininio mokymosi metodai – ra grindžiami gebėjimu atpažinti įvairias savybes mokymosi duomenų aibėje, neturint jokių nurodymų, kokių rezultatų tikimasi [21]. Atlikus mokymosi duomenų analizę, šie metodai gali atlikti atitinkamas prognozes apie naujus, anksčiau nematytus duomenis. Neprižiūravimo mašininio mokymosi metodai dažnai pritaikomi grupavimo (angl. *clustering*) bei dimensijų mažinimo (angl. *dimensionality reduction*) uždaviniuose [21]. Klasterizavimui dažnai naudojami k-vidurkių ir k-medoidų metodai, o pagrindinių komponentų analizė (angl. *principal component analysis*, PCA) yra vienas iš taikomų dimensijų mažinimo metodų pavyzdžių. Šių metodų pagrindinis skirtumas nuo prižiūravimo tipo algoritmų yra tas, kad duomenys neturi turėti jiems iš anksto priskiriamų kategorijų ar numatytų rezultatų. Tai leidžia sumažinti reikalingą žmogiškųjų išteklių darbą ir sutrumpina duomenų rinkimo bei apdorojimo etapus.

Neprižiūravimo metodų taikymas sentimentų analizės užduotims yra efektyvus, nes sentimentų žodžiai ir jų grupės tekste dažnai atlieka dominuojantį vaidmenį nuomonių gavybos procese. Šie metodai naudoja iš anksto nustatytus algoritmus tam, kad atpažintų tam tikrus šablonus dokumentuose, kurie gali būti naudingi emocijų nustatymui. Šiuose metoduose gali būti pritaikomi sintaksės pagrindu sudaromi šablonai, sudaryti remiantis POS žymėjimais [12]. Norint taikyti šiuos metodus sentimentų analizei yra taikoma keletas veiksmų seka bei įvairių metodų panaudojimas. Vienas iš gana dažnai taikomų metodų pirmiausia remiasi sentimentų žodyno (angl. *sentiment lexicon*) sukūrimu jo neprižiūrint ir tada pritaikant tam tikrą funkciją, kuri gebėtų nustatyti pozityvumo (ar subjektyvumo) laipsnį pasirinktoms teksto grupėms remiantis apibrėžtu žodynu. Vieni pirmųjų tokius mokymosi būdus pritaikė V. Hatzivassiloglou, J. Wiebe bei P. Turney [6]. Esminis šių metodų komponentas yra technika, kuria pasirenkama sudaryti sentimentų žodyną bei su jais siejamus sentimentų poliariškumo rodiklius. Vienas iš galimų pritaikymo metodų yra lingvistikos euristikos (angl. *linguistic heuristics*) nagrinėjimas. Šie metodai grindžiami poliariškumo klasifikavimu, kai duomenys grupuojami į viena kitai priešingas reikšmes turinčias kategorijas. Apmokius metodus, duomenims priskiriamos teigiamos arba neigiamos orientacijos kategorijos. Šios kategorijos gali būti priskiriamos atsitiktinai, remiantis dažniausiai pasikartojančiomis reikšmėmis,

arba naudojant papildomus žodynus, pavyzdžiui, *WordNet*, kuriame aprašomi žodžių tarpusavio ryšiai ir jiems priskiriami sentimentai. Alternatyvus klasifikavimo būdas buvo pasiūlytas P. Turney, kuris yra pritaikytas teigiamais bei neigiamais ‘nykščiais’⁷ paremtų vertinimų semantinei orientacijai nustatyti ir iki šių dienų dažnai taikomas praktikoje [6]. Dar kitas taikomas metodas yra neprižiūrimo mokymo metodu gautų suklasifikuotų rezultatų panaudojimas prižiūrimojo mokymo algoritmams apmokyti. Buvo pastebėta, kad šio metodo modifikacijos efektyviai atpažįsta subjektyvius dokumentus [6]. Apibendrinant, neprižiūrimieji mašininio mokymosi metodai dažnai taikomi kaip pagalbiniai sentimentų analizės uždaviniams spręsti. Šių metodų privalumas yra gebėjimas atskleisti įvairius požymius analizuojamame tekste ir panaudoti juos tolimesniems kategorijų priskyrimo etapams.

1.3.2.3. Giliuoju mokymusi pagrįsti metodai

Giliuoju mokymusi pagrįsti metodai (angl. *deep learning approach*) – tai mašininio mokymosi poaibis. Šie suteikia galimybę kompiuteriams mokytis iš savo patirties ir suprasti pateiktus duomenis kaip tam tikrų idėjų hierarchiją (žr. 1.3.1 pav.) [24]. Kiekviena aptinkama sąvoka/požymis gali būti apibrėžiamas dar paprastesnėms sąvokomis ar detalėmis. Gebėdami kaupti žinias iš patirties, šie metodai nereikalauja žmonių aprašyti visų reikiamų instrukcijų kaip atlikti tam tikrus uždavinius. Sudarant aptinkamų sąvokų ar požymių hierarchijas, šie metodai suteikia galimybę atpažinti ir suprasti sudėtingas idėjas, kurias sudaro smulkūs, aiškiai apibrėžti elementai [24]. Šie metodai ypač išpopuliarėjo sprendžiant balso atpažinimo, vaizdų identifikavimo bei NLP problemas.



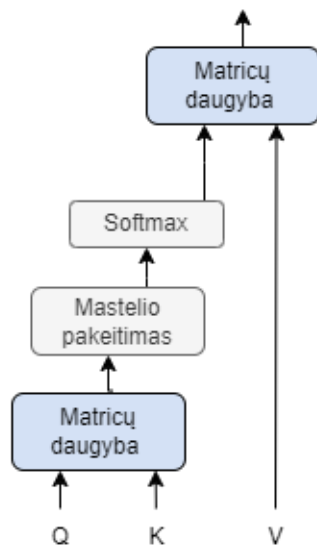
1.3.1 pav. Giliuoju mokymusi paremtų mašininio modelių struktūros išskirtinumas (sudaryta remiantis [25])

Nuomonių gavybos uždavinys gali būti apibrėžiamas kaip požymių lygio (angl. *token-level*) kategorizavimo uždavinys. 2010-aisiais – 2018-ais metais atlikti tyrimai bandė sukurti tekstų požymių nusakymui pritaikytus vektorius, kurie pasižymėtų nedideliu dimensionalumu, tankiu bei pritaikomu turiniu. Buvo pradėti naudoti sąlyginių atsitiktinių laukų (angl. *conditional random field*, CRF) ir jų variantai, kaip daliniai CRF, statistiniai modeliavimo metodai. CRF modeliai derinami su rankiniu būdu apibrėžtais žodžių, frazių ir sintaksės požymiais, todėl jų tikslumas priklauso nuo apibrėžtų požymių rinkinių kokybės. Šie rinkiniai dažniausiai sudaromi remiantis sąryšių medžiais ar rankiniu būdu sukurtomis žodynais [26]. CRF algoritmų rezultatai buvo pagerinti pasitelkiant pasikartojančių neuroninių tinklų (angl. *recurrent neural network*, RNN) metodus. RNN yra grįžtamojo neuroninio tinklo (angl. *feedforward neural networks*, FNN) metodų plėtinys. RNN modeliai išsiskyrė savo gebėjimu apdoroti kintamo dydžio sekas ir tapo ypač populiarūs NLP uždaviniuose [17]. RNN metodai, turintys susietus daugybinius sluoksnius, sugeba pritaikyti laiko

⁷ Tekstų grupavimas turint priskiriamą kategoriją bendram įvertinimui remiantis *nykščio* indikatoriais (angl. *thumbs-up /thumbs-down*)

skalėje kintančią hierarchiją: žemesniame lygyje esantys sluoksniai atpažįsta trumpalaikius ryšius tarp žodžių, o aukštesniuose lygiuose aptinkamos interpretacijos tarp tolesnių ir ilgesnių tekstų iškarpu. [26]. Tačiau RNN modeliai susidūrė su nykstančio gradiento problema bei negebėjimu įsiminti ilgesnio teksto konteksto. Šias problemas spręsti padėjo patobulinti ilgos trumpalaikės atminties (angl. *long short-term memory* LSTM) modeliai.

Reikšmingas proveržis NLP uždavinių srityje įvyko 2014 metais, kai buvo paskelbtas straipsnis „Sekų mokymasis su neuroniniais tinklais“ (angl. „Sequence to Sequence Learning with Neural Networks“) [27]. Straipsnio autoriai pristatė modelį, sudarytą iš keleto LSTM sluoksnių *koduotuvo–iškoduotuvo* (angl. *encoder–decoder*) architektūroje [27]. Sukurto modelio rezultatai paspartino tolimesnius tyrimus. 2017 metais buvo pristatytas svarbus straipsnis apie dėmesio mechanizmų (angl. *attention mechanism*) taikymą mašininio mokymosi modeliuose, grindžiamuose giliuoju mokymu. *Google Brain* ir Toronto universiteto mokslininkai straipsnyje „Dėmesys – viskas, ko jums reikia“ (angl. „Attention is all you need“) pristatė naują tinklo architektūrą – transformerių modelius [28]. Pristatytas modelis patobulino LSTM koduotuvo–iškoduotuvo architektūrą, suteikiant galimybę modeliui atsižvelgti į gaunamos duomenų įvesties kontekstą. Modelis buvo pagrįstas dėmesio mechanizmu atsisakant pasikartojimų ir konvoliucinių skaičiavimų [28]. Publikacijoje pristatomas pirmasis sekų transdukcijos (angl. *sequence transduction*) modelis, kuris rekurentinius sluoksnius pakeičia daugiagalviu dėmesio (angl. *multi-headed attention*) mechanizmu koduotuvų–iškoduotuvų architektūrose. 1.3.2 pav. pateiktas autorių pristatyta skaliarinės sandaugos dėmesio mechanizmo architektūra.



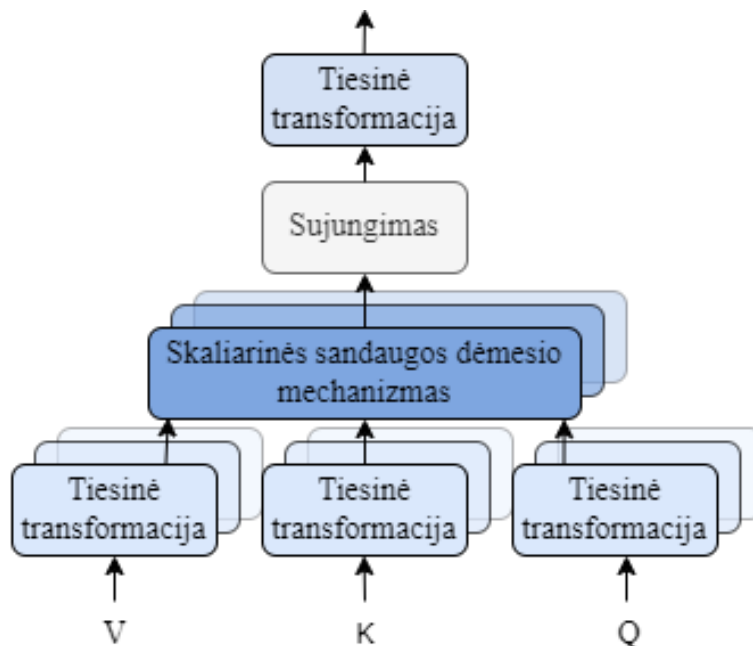
1.3.2 pav. Sumažintas skaliarinės sandaugos dėmesio mechanizmas (sudaryta remiantis [28])

Siekiant apdoroti ilgesnės apimties tekstus susiduriama su taikomų vektorių mažo dimensionalumo problema. Dėl šios priežasties, modeliai turi ribotą aktyvaus įsimenamo konteksto ilgį. Dėl šios priežasties buvo pristatyta galimybė atsižvelgti tiesiogiai į visą įvesties tekstą kai yra bandoma atlikti prognozavimą. 1.3.2 pav. pavaizduotas dėmesio įverčio apskaičiavimas yra išreiškiamas (3) formule [28].

$$\text{Dėmesio\textit{vertis}}(Q, K, V) = \textit{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V; \quad (3)$$

čia:

- Q – tai vadinama užklauso (angl. *Query*) matrica, kuri atstoja pagrindinę žymą (išskirtą žymų išskyrimo metu), kuriai siekiama nustatyti reikšmingą kontekstą;
- K – tai vadinama raktų (angl. *Key*) matrica, skirta sukurti raktų vektorius visoms įvesties žymoms. Šie vektoriai leidžia algoritmui įvertinti sąryšį ar susietumą (angl. *relevance*) tarp pagrindinio žodžio ir kitų žodžių sakinyje. Kuo didesnis įverčio balas, tuo stipresnis numatomas ryšys tarp nagrinėjamų žodžių;
- V – tai vadinama vertės (angl. *Value*) matrica, kuri sukuria reikšmių vektorius visoms įvesties žymoms. Šie vektoriai savyje saugo informaciją apie kiekvieno žodžio kontekstą, kurių galutinės reikšmės atsižvelgia į labiausiai svarbius ir susijusius žodžius;
- d_k – raktų vektoriaus dimensija, skirtas koreguoti matricų daugybos rezultatus, siekiant išvengti per didelių gradientų reikšmių skaičiuojant *softmax* funkciją.



1.3.3 pav. Daugiagalvis dėmesio mechanizmas (sudaryta remiantis [28])

1.3.3 pav. pateikiamas straipsnyje pristatytas naujas apmokymo būdas, pritaikant daugiagalvį dėmesio mechanizmą (angl. *multi-head attention*). Vietoje atliekant vieną dėmesio sutelkimo funkciją, buvo pristatyta galimybė tiesiškai projektuoti (angl. *linearly project*) užklauso h kartų, pasirenkant ir atitinkamai naudojant skirtingas gautas tiesines projekcijas į d_V , d_K ir d_Q dimensijas. Kiekvienas šis projektuojamas etapas tada gali apskaičiuoti dėmesio funkcijos reikšmes lygiagrečiai ir grąžinti d_V dimensijos reikšmę. Šis dimensionalumas suteikia galimybę modeliui suvokti įvairias duomenų charakteristikas. Visos šios reikšmės sujungiamos ir toliau projektuojamos kaip pateikiama nurodytoje diagramoje [28]. Daugiagalvis dėmesio mechanizmas suteikia galimybę įgyti platesnį dėmesio sutelkimo spektrą, leidžia skirti dėmesį informacijai iš įvairių reprezentacinių pogrupių skirtingose pozicijose. Tai užtikrina lankstumą ir gebėjimą atrasti įvairaus tipo požymius, kurie būtų neatpažinti, jei būtų naudojama tik viena dėmesio atšaka.

Dėmesio mechanizmu paremta modelio architektūra buvo pritaikyta sukurti didelės kalbos modelius (angl. *large language models, LLM*). Šis modelių pritaikymas yra ypatingas tuo, kad iš pradžių modeliai mokomi naudojantis dideliu duomenų kiekiu, jų neklasifikuojant ar kitaip neapdorojant ir taikant neprižiūrimojo tipo mokymąsi. Taip modeliai įgauna bendros paskirties galimybes bei žinių

apie kalbą, kurios gali būti perduotos tolimesnėms užduotims spręsti [29]. Šie modeliai vėliau pritaikomi įvairiems konkrečių užduočių sprendimams, tokioms kaip tekstų vertimas, žodžių klasifikavimas, pokalbių imitavimas ar sentimentų analizė. Šiandien žinoma nemažai alternatyvių sentimentų analizės LLM modelių pritaikytų anglų kalbai, kurie pateikia gana aukštus rezultatus apibrėžtuose duomenų imtyse. Tačiau vis dar nėra gerai žinomų tyrimų, pasiekiančių aukštus rezultatus, taikant daugiakalbę arba mažiau populiarios kalbos, kaip lietuvių, sentimentų analizę.

1.3.2.4. Mašininio mokymusi paremtų metodų apibendrinimas

Apibendrinant, šiandien pristatoma daugybė įvairių mokslinių straipsnių, aprašančių giliojo mašininio mokymosi taikymo galimybes sentimentų analizės problemoms spręsti. Nors gilusis mokymas moksliniuose tyrimuose demonstruoja puikius rezultatus atpažįstant nuotraukas ir balsą, jis nesukėlė reikšmingo proveržio NLP srityje [17]. Naujausios tendencijos kalbos apdorojimo užduotyse atsirado pristatant didelės kalbos modelius, kurie suteikė gerokai geresnius rezultatus, palyginti su anksčiau naudotais RNN modeliais. Vis dėlto iki šiol tam tikri uždaviniai toliau analizuojami įvairių mokslininkų, siekiant gerinti esamus rezultatus. Tai rodo, kad tekstinių duomenų apdorojimo uždaviniai išlieka sudėtingi, o taikomi metodai gali būti tobulinami ir derinami su kitomis duomenų analizės strategijomis, siekiant pagerinti sentimentų analizės rezultatus.

1.4. Sentimentų analizėje taikomų metodų apžvalga

Šiame skyriuje aptariami darbo metu nagrinėjami ir praktikoje taikomi natūralios kalbos apdorojimo metodai, mašininio mokymosi algoritmai, detalizuojami didelių kalbos modelių veikimo principai bei kiti susiję duomenų analizavimo ir apdorojimo metodai. Detaliau aprašomi transformerio tipo kalbos modeliai ir jiems būtini duomenų apdorojimo procesai, kaip žodžių žetonizacija (angl. *tokenization*), vektorizacija (angl. *vectorization*) ir kiti apdorojimo procesai.

1.4.1. Teksto žetonizacija

Žetonizacija – tai teksto apdorojimo etapas, kurio metu gauti tekstiniai duomenys yra suskaidomi į žodžius ar trumpas frazes vadinamas žetonais (angl. *tokens*) [30], kartais žinomas kaip žymas. Žetonai gali padėti bandant suprasti duomenų kontekstą ir kuriant modelius NLP uždavinių sprendimui [30]. Šie metodai taikomi siekiant struktūrizuoti duomenis taip, kad modeliai galėtų juos suprasti. Žetonizacija yra vienas pirmųjų ir svarbiausių žingsnių natūralios kalbos apdorojimo išankstinio paruošimo etape [31]. Paprasčiausias šio metodo pavyzdys yra sakinio „Šiandien šviečia saulė“ išskaidymas į žymas: „Šiandien“, „šviečia“, „saulė“. Vieni žymiausių ir dažniausiai taikomų metodų yra šie [30]:

- *tarpais paremta žetonizacija* (angl. *white space tokenization*) – tai metodas, kada gauti duomenų tekstai yra išskaidomi pagal aptinkamus tarpo skyrybos ženklus;
- *žodynu paremta žetonizacija* (angl. *dictionary based tokenization*) – tai metodas, kai žetonai išskiriami pagal egzistuojančius žymų aprašus pasirinktuose žodynuose. Tekstai, kuriuose nėra aptinkama žinomų žetonų, turi būti apdorojami pagal kitas taisykles;
- *taisyklėmis paremta žetonizacija* (angl. *rule based tokenization*) – tai metodas, kai žodžiai suskaidomi pagal iš anksto apibrėžtas taisykles (pvz.: gramatikos taisyklės);
- *įprastomis išraiškomis paremta žetonizacija* (angl. *regular expression tokenization*) – tai metodas, kuris paremtas įvairių išraiškų pritaikymui teksto žetonų išskyrimo uždavinyje;
- *dalinių žodžių žetonizacija* (angl. *subword based tokenization*) – tai metodas, kuris naudojamas priskirti unikalius identifikatorius žodžių dalims. Šie žetonai yra sudaromi

remiantis žodžių ir jų dalių pasikartojimo dažnumu. Rečiau naudojami ar sudėtingesni žodžiai skaidomi į dalis, kurios geriau atspindi prasmę. Pavyzdžiui, žodis ‘*puikiausias*‘ galėtų būti išskirtas į dalis ‘*puik*‘ ir ‘*iausias*‘, taip suteikiant galimybę modeliui išmokti pagrindinę žodžio prasmę. Šia metodikai galima naudoti įvairius algoritmus, kaip:

- dvejetainių porų kodavimas (angl. *byte-pair encoding*, BPE);
- žodžio dalies (angl. *word-piece*) išskyrimas;
- unigramų kalbos modelis.

Tai tik keletas pavyzdžių iš žinomų ir taikomų žetonizacijos metodų. Jau yra sukurti ir pritaikyti modeliai anglų kalbos tekstų žetonams išskirti. Tačiau praktikoje, ypač naudojant didelio masto kalbos modelius, dažniausiai sutinkamas dalinių žodžių žetonizacijos metodas.

1.4.2. Teksto vektorizavimo metodai

Vektorizavimas – tai procesas, kurio metu teksto dokumentų rinkiniai yra pakeičiami į skaitmeninių ypatybių vektorius [32]. Ši strategija (žetonų išskyrimas, reikšmių perskaičiavimas ir normalizavimas) suteikia galimybę mašiniams modeliams suprasti tekstą išreikštą skaitinėmis reikšmėmis. Vektorizacija suteikia galimybę aprašyti dokumentus pagal juose esančių žodžių pasikartojimų kiekį nevertinant žodžių pozicijos bei sąryšių su kitais žodžiais esamame tekste. Šis uždavinys sudėtingas dėl žodžių įvairovės bei įprastų žodžių pasikartojimo dažnumo [32]. Dažnai tenka pasirūpinti aptiktų požymių dimensijų sumažinimu bei labiausiai pasikartojančių, reikšmės nekeičiančių žodžių pašalinimu ar mažesnio svorio vertinimu. Praktikoje yra taikomi įvairūs teksto vektorizavimo metodai, kaip žodžių rinkinio krepšelio, žodžių maišos metodai ir kiti. LLM metodai dažniausiai naudoja žodžių maišos, pozicijų ir segmentų įterpimo bei konvertavimo metodus, kurie sutelkia dėmesį į semantinių ryšių tarp žodžių identifikavimą ir užfiksavimą.

1.4.2.1. Žodžių rinkinio krepšelis

Žodžių rinkinio krepšelio metodas (angl. *bag of words*, BoW) – tai yra paprasčiausias metodas taikomas teksto pavertimui į skaitmenines reikšmes [33]. Šis algoritmas neįvertina žodžių tvarkos sakiniuose, o kiekvienas gaunamas žodis savaime yra suprantamas kaip požymis [33]. Šis modelis yra pritaikytas sudaryti fiksuoto dydžio žodžių požymių vektorius.

BoW metodas yra paremtas naudojamo žodyno modelio sudarymu. Žodynas gali būti sudaromas remiantis visais turimais ir analizuojamais unikaliais žodžiais. Šie žodžiai įgauna atitinkamas reikšmes žodyne. Analizuojamas tekstas tada suskaidomas individualiais žodžiais ir yra apskaičiuojamas kiekvieno žodžio pasikartojimų dažnis tekste [33]. Pavyzdžiui, jei analizuosime sakinius: ‘*Šiandien labai puikus rytas!*‘ bei ‘*Puikus rytas ir labai puikus vakaras*‘, mūsų žodyną sudarytų 7 skirtingi žodžiai. Jus apdorojant BoW metodu sudaroma 1.4.1 lentelė.

1.4.1 lentelė. BoW metodų paremtas žodžių pasikartojimų išskyrimo pavyzdys

	1	2	3	4	5	6	7	Sakinio ilgis žodžiais
	šiandien	labai	puikus	rytas	ir	labai	vakaras	
1-asis sakiny	1	1	1	1	0	0	0	4
2-asis sakiny	0	1	2	1	1	1	1	7

Remiantis šia lentele sakinius sudarantys vektoriai atitinkamai būtų atvaizduojami kaip [1111000] bei [0121111]. Siekiant šį metodą padaryti kuo tikslesnį, analizuojamuose tekstuose dažnai yra

suvienodinamos mažosios bei didžiosios raidės, pasirenkami tik žodžių kamienai ir panašiai. Vis dėlto, šie metodai yra neatsparūs naujiems terminų įtraukimams, kadangi pridėjus naujus žodžius žodynuose – pasikeis sukuriamas žodžių vektorius ilgis. Jei nusprendžiama ignoruoti neatpažįstamus žodžius, didėja tikimybė, kad bus prarandama svarbi informacija.

BoW metodo taikymas sentimentų analizės uždaviniuose yra dažnai sutinkama praktika. Tačiau atliekant palyginamuosius mokslinius tyrimus yra pastebima, kad remiantis vien tik BoW metodu rezultatai nėra patys geriausi. Šis metodas gali būti pritaikomas įvairių sentimentų vertinimui išskiriant daugiau nei keletą kategorijų. Tačiau, galiausiai šis modelis praranda daug reikšmingos informacijos, kuri padeda nustatyti tekste perteikiamus sentimentus [5].

1.4.2.2. Termino dažnis – atvirkštinis dokumento dažnis

Termino dažnis – atvirkštinis dokumento dažnis (angl. *term frequency-inverse document frequency*, TF-IDF) – tai vienas populiariausių žodžių svorio priskyrimo algoritmų, kuris priskiria tekste aptinkamiems terminams jų svarbos skaitinę reikšmę analizuojamo dokumento kontekste [34]. Šis algoritmas nustato žodžio svarbą pagal jo pasikartojimo dažnį ir atitinkamą dokumento reikšmę kontekste [34]. Taikant šį metodą yra atliekami trys skaičiavimo žingsniai [33]. Pirmiausia yra nusakomas termino dažnis naudojantis (4) formule.

$$tf_{t,d} = \frac{n_t}{f(t,d)}; \quad (4)$$

čia $f(t,d)$ yra termino t pasikartojimų kiekis dokumente d . Šis skaičiavimas yra atliekamas siekiant nustatyti žodžio pasikartojimo dažnį (angl. *term frequency*, TF). Toliau remiantis bendru dokumentų kiekiu N bei visų turimų dokumentų D , kuriuose yra aptiktas terminas w , kiekiu yra apskaičiuojamas atvirkštinis dokumento dažnis idf_t žodžiui t remiantis (5) formule.

$$idf_{t,D} = \log\left(\frac{N}{f(t,D)}\right) \quad (5)$$

Galiausiai, turint TF bei IDF reikšmes galima nusakyti termino TF-IDF pritaikant (6) formulę:

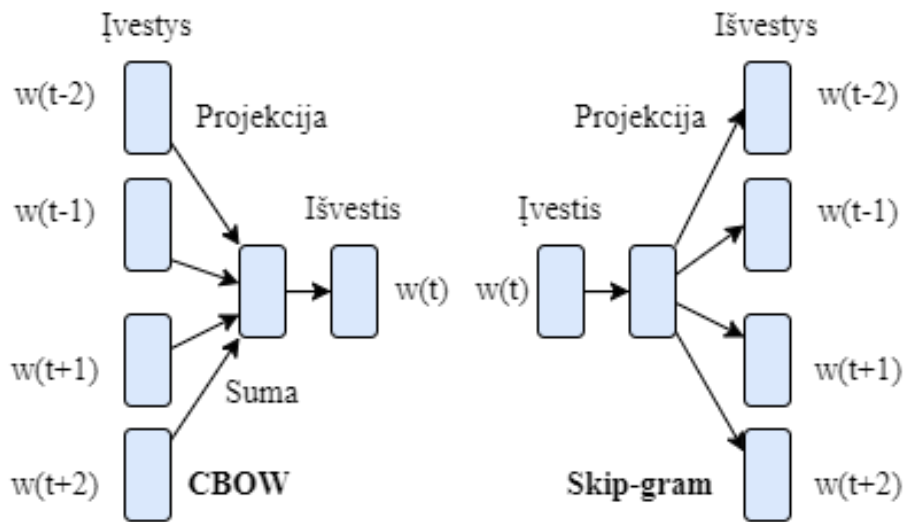
$$tfidf_{t,d,D} = tf_{t,d} \cdot idf_{t,D}. \quad (6)$$

TF-IDF metodo galutinis rodiklis suteikia didesnes svorių reikšmes mažiau pasikartojantiems žodžiams. Aukšta šio įverčio reikšmė parodo, kad analizuojamas terminas yra retai sutinkamas visų dokumentų rinkinyje, tačiau dažnai pasikartojantis analizuojamame tekste [33]. Šis modelis gali perteikti daugiau informacijos apie svarbesnius ir mažiau svarbius žodžius aptinkamus tekstų rinkiniuose.

1.4.2.3. Word2Vec

Word2Vec – tai žodžių vektorizavimo modelis, pirmą kartą pristatytas 2013 metais [35]. Šis geba atpažinti sintaksės ir semantinius panašumus tarp žodžių. Jo architektūra yra paprastesnė lyginant su neuroniniais tinklais paremtais modeliais. Tačiau dėl mažesnių skaičiavimo techninių iššūkių, modelis geba sukurti didelio dimensionalumo žodžių vektorius iš didelės apimties duomenų [35]. Tai leidžia sukurti modelius iš be galo didelių žodynų. *Word2Vec* apima kelias skirtingas modelių architektūras ir optimizacijas, kurios leidžia atpažinti ir išskirti žodžių reikšmes iš didelių duomenų

imčių. Dvi pagrindinės naudojamos architektūros yra tęstinas žodžių rinkinio krepšelio (angl. *continuous bag of words*, CBOW) bei tęstinas praleidimo-gramos (angl. *continuous skip-gram*) modelis.



1.4.1 pav. CBOW ir Skip-gram modelių architektūros (sudaryta remiantis [35])

1.4.1 pav. pateiktos *Word2Vec* algoritmo modelių architektūros, kurios yra naudojamos sudaryti pasiskirstyta žodžių atvaizdavimą. CBOW struktūroje projekcijos sluoksnis yra bendras visiems žodžiams (o jų vektorių reikšmėms apskaičiuojami vidurkiai). Žodžių eiliškumas nedaro įtakos suformuojamai projekcijai. Be to prognozuojamam žodžiui naudojami prieš jį ir po jo einančių žodžių rinkiniai. *Skip-gram* architektūra, nors ir panaši į CBOW, yra pritaikyta atlikti atvirktinį funkcionalumą. Pateikus įvesties žodį $w(t)$, modelis prognozuoja žodžius prieš ir po dabartinio žodžio. Dėl praleidžiamos prognozės $w(t)$, modelis vadinamas praleidimo-grama. Kadangi toliau esantys žodžiai yra mažiau susiję su pagrindiniu įvesties žodžiu, jiems priskiriama mažesnė svorio reikšmė. Šių algoritmų kombinacija, apmokant modelį su pakankamai dideliu duomenų kiekiu, suteikia galimybę atvaizduoti žodžių kontekstą apibrėžtose n -dimensijose. Tai leidžia aptikti įvairaus pobūdžio panašumus tarp žodžių.

$$X = \text{vektorius}(\text{didžiausias}) - \text{vektorius}(\text{didelis}) + \text{vektorius}(\text{mažiausias}) \quad (7)$$

Ilgautos vektorių reikšmės, leidžia atlikti paprastas algebros operacijas su žodžių reprezentacijomis [35]. Pavyzdžiui, pateiktoje (7) formulėje, bandoma aptikti žodį X panašų į *mažas* perteikiant tą pačią prasmę, kaip žodis *didelis* yra panašus į žodį *didžiausias*. Šio tipo algoritmas pasižymėjo aukštais rezultatais siekiant nustatyti susietų žodžių prasmę ir reikšmes. Tačiau svarbu pabrėžti, kad tai algoritmas, kuris sukuria vieną unikalų vektorių atitinkamam žodžiui, todėl tokio tipo modeliai turi ribotas galimybes bendrojo pobūdžio kalbos uždaviniuose.

1.4.2.4. Žodžių įterpimai

Kontekstų paremti žodžių įterpimo (angl. *word embeddins*) metodai – tai teksto vektorizavimo metodai, kurie atvaizduoja žodžius fiksuoto ilgio vektoriais. Šiuose methoduose naudojamas skaitinis atvaizdavimas leidžia identifikuoti ir užfiksuoti lingvistines kalbos savybes, kaip morfologija, sintaksė, leksika [36]. Žodžių įterpimo metodai pasižymi gebėjimu panašiams žodžiams suteikti

skirtingus apibrėžimus ir savybes priklausomai nuo esamo konteksto, kuriame jie įterpiami. Tradiciniai žodžių įterpimo metodai buvo sukurti apmokant modelius universalumu bei abstraktumu paremtais ryšiais. Šie metodai yra paremti arba visiškai sudaryti iš žodynų atrenkant atitinkamus unikalius žodžius. Tačiau remiantis tokiais principais, sudėtinga apmokyti aukštos kokybės ryšius. Šie metodai idealiu atveju turėtų sugebėti identifikuoti sudėtingas žodžių naudojimo charakteristikas (pvz. semantika, sintaksė) bei atpažinti žodžių polisemiją⁸ [37]. Naujesni įterpimo metodai siekia atpažinti ir aptikti žodžių kontekstą, taip sprendžiant minėtas problemas. Tokio tipo metodai yra patempti neuroninių tinklų modelių gebėjimu sukurti žodžių žymų vektorius užfiksuojančius teksto prasmę. Konteksto įvertinimas leidžia vizualiai atvaizduoti žodžius atspindint jų dinaminę reikšmę, o tai reiškia, kad vienas žodis gali turėti kelis skirtingus vektorius [38].

2018 metais buvo pristatytas pirmasis savarankiško mokymosi žodžių įterpimų kūrimo didelės kalbos modelis ELMo (angl. *embeddings from language models*) [37]. Šiame modelyje žodžių atvaizdavimas yra gražinamas kaip funkcijos rezultatas, priklausantis nuo viso įvesties sakinio. Žodžių įterpimai yra apskaičiuojami remiantis dviejų lygių dvikrypčių modelių pagrindu biLMs. ELMo modelis pagrįstas dvikrypte LSTM neuroninio tinklo struktūra [37]. Tokia struktūra suteikia galimybę atlikti dalinai prižiūrimo mokymąsi ir šią vektorizavimo struktūrą įtraukti į didesnes modelių architektūras. Po šio modelio buvo sukurta daugiau įvairių, kontekstą atpažįstančių, žodžių įterpimo metodų. Jie tapo labai populiarūs apdorojant tekstus didelės kalbos modelių architektūrose. Dėl jų gebėjimo veikti dinamiškai ir įvertinti kalbos semantines bei sintaksines ypatybes, įterpimais paremti modeliai žymiai pagerino modelių tikslumą, ypač atliekant tekstų klasifikavimo uždavinius.

1.5. Didieji kalbos modeliai taikomi sentimentų analizėje

Mašininio mokymosi metodai sentimentų analizėje buvo pradėti taikyti dar 2000-ųjų metų pirmojo dešimtmečio pradžioje. „*Pirštas aukštyn?*“ (angl. „*Thumbs up?*“) buvo pirmoji plačiai dėmesio sulaukusi publikacija, skirta nuomonių gavybos uždaviniui, kuri buvo pristatyta dar 2002 metais [11]. Šis straipsnis sulaukė didelio susidomėjimo ir nulėmė tolimesnių tyrimų kryptį. Iš pradžių sentimentų klasifikavimui buvo taikomi įvairūs prižiūrimo tipo mašininio mokymosi klasifikavimo metodai, tokie kaip sprendimų medžiai, atraminių vektorių mašinos ar naivusis Bajeso klasifikavimo algoritmas. Tačiau, proveržis giliojo mašininio mokymosi srityje pritaikant transformerių modelius, leido pasiekti pastebimai geresnių rezultatų kalbos uždaviniuose. Pritaikytas išankstinis mokymasis dideliems modeliams pasirodė ypač efektyvus gerinant rezultatus NLP uždaviniuose. Šiuo metu geriausiai žinomos dvi strategijos, kaip panaudoti apmokytus kalbos modelius tolimesniems tyrimams [39]:

- savybių išskyrimas (angl. *feature-based approach*) – naudoja architektūras pritaikytas specifinei užduočiai. Šios architektūros yra kuriamos taip, kad naudotų jau apmokytas dalines žodynų reprezentacijas (pvz.: ELMo);
- modelio koregavimas (angl. *fine-tuning*) – suteikia galimybę įvesti naujus, specifinei užduočiai pritaikytus parametrus, kurie tada naudojami tolimesnėms apmokymo užduotims.

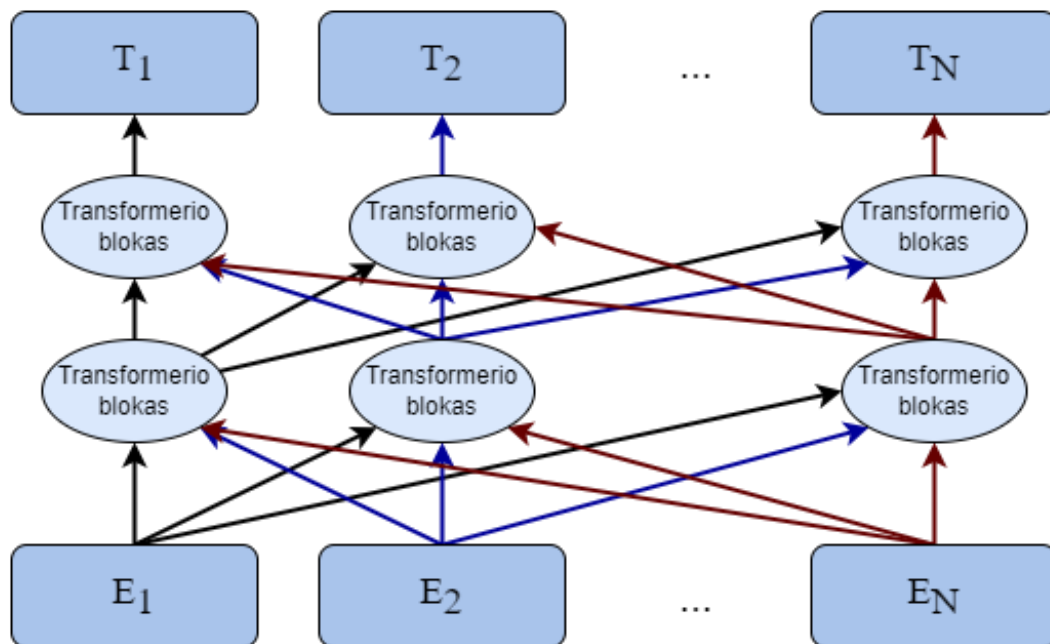
Šie metodai remiasi to pačio principo funkcijomis išankstinio mokymo metu, naudojant vienakrypčius kalbos modelius bendrai kalbos reprezentacijai mokant. Šio skyriaus ribose toliau išsamiai detalizuojamos ir aptiriamos dvi transformerių modelių architektūros, kurios gali būti pritaikytos sentimentų analizės uždaviniams spręsti.

⁸ Polisemija – žodžio įvairių reikšmių turėjimas priklausomai nuo konteksto, daugiareikšmiškumas.

1.5.1. BERT tipo modeliai

Dvikryptis transformacinio kodavimo būdas (angl. *Bidirectional Encoder Representations from Transformers*, BERT) – tai pažangi ir plačiai žinoma transformerio modelio architektūra, pristatyta 2018 metais [39]. Esminis pristatytas modelio išskirtinumas tai jo gebėjimas mokytis giliausias dvipuses reprezentacijas iš įvesties duomenų, vienodai atsižvelgiant į aplinkinių žodžių kontekstus (žodžius einančius prieš ir po analizuojamo žodžio sakinyje) visuose sluoksniuose [39]. Dėl to, iš anksto apmokytas BERT modelis gali būti koreguojamas vos su vienu papildomu išvesties sluoksniu ir taip būti pritaikomas plačiam kalbos užduočių spektrui.

1.5.1.1. Modelio struktūra



1.5.1 pav. BERT dvikryptė apmokymo modelio architektūra (sudaryta remiantis [39])

BERT modelio architektūra tai daugiasluksnė dvikryptė koduotuvo–iškoduotuvo transformerio architektūra (žr. 1.5.1 pav.). Modelio kuriamos reprezentacijos yra derinamos iš visų krypčių, kiekviename sluoksnyje. Transformerių modelių architektūroje, koduotuvo dalyje naudojama savaiminio dėmesio (angl. *self-attention*) struktūra, kuri taip pat išlaikyta BERT architektūroje. Pradinė pristatyta architektūra iki šių dienų sulaukė įvairių modelio modifikacijų, kurios apima skirtingą modelio parametrų kiekį, skirtingas apmokymo duomenų imtis ar kitus parametrus. Originaliame BERT architektūros pristatyme, buvo pateiktos dvi pagrindinės architektūros: BERT_{BASE} ir BERT_{LARGE}. Svarbiausi parametrai aprašantys modelio dydį yra:

- L – tai sluoksnių (kitais žinomu, kaip transformerio bloku) skaičius;
- H – tai paslėptų sluoksnių skaičius;
- A – tai savaiminio dėmesio galvų skaičius.

Pagal šiuos įverčius BERT_{BASE} iš viso turi 110 milijonų parametrų, kai $L = 12$, $H = 768$, $A = 12$, o BERT_{LARGE} modelis turi 340 milijonų parametrų, $L = 24$, $H = 1024$, $A = 16$.

1.5.1.2. Žetonizacija

BERT modelis žetonų išskyrimui naudoja *WordPiece* algoritmą. BERT modelyje naudojamą *WordPiece* žodyną sudaro 30,000 žetonų. Išskirtinės taikomos savybės algoritmui [39]:

- kiekviena įvestis visada yra pradedama klasifikavimo ženklu *[CLS]*;
- galutinė paslėpta būseną, atitinkanti *[CLS]* ženklą, naudojama kaip bendra sekos reprezentacija klasifikavimo užduotims;
- jei sakinių poros sujungiamos į vieną seką, algoritmas atskiria sakinį dviem būdais: naudojant *[SEP]* ženklą ir segmente nurodant ar jis priklauso sakiniui A ar B.

WordPiece modelis yra žodžių dalių išskyrimo pagrindu veikiantis žetonizavimo metodas. Šis algoritmas užtikrina deterministinį segmentavimą bet kuriai galimai apmokymo įvesties sekai [40]. Žodžių dalių pagrindu veikiantys žetonizatoriai, pirmiausia tekstą skaido pagal žodžių segmentus. Dėl šios priežasties tarpai tarp žodžių yra panaikinami ir žetonizavimo procesas yra negrįžtamas [41]. Nors *Google* nėra atskleidusi *WordPiece* algoritmo detalių, yra pateikti apibendrinti veiksmi. Metodas sudaro žetonų išskyrimo žodyną iteraciniu būdu. *WordPiece* pradeda nuo mažos žodyno apimties ir specialiųjų ženklų derinio. Toliau atsižvelgdamas į žodžių pasikartojimo dažnumą ir kombinacijas, algoritmas iteratyviai sujungia dažniausiai pasitaikančias žodžių dalis. Vietoj tiesioginio dažniausiai pasitaikančios poros pasirinkimo, *WordPiece* skaičiuoja kiekvienos poros įvertį, remdamasis (8) formulę [42].

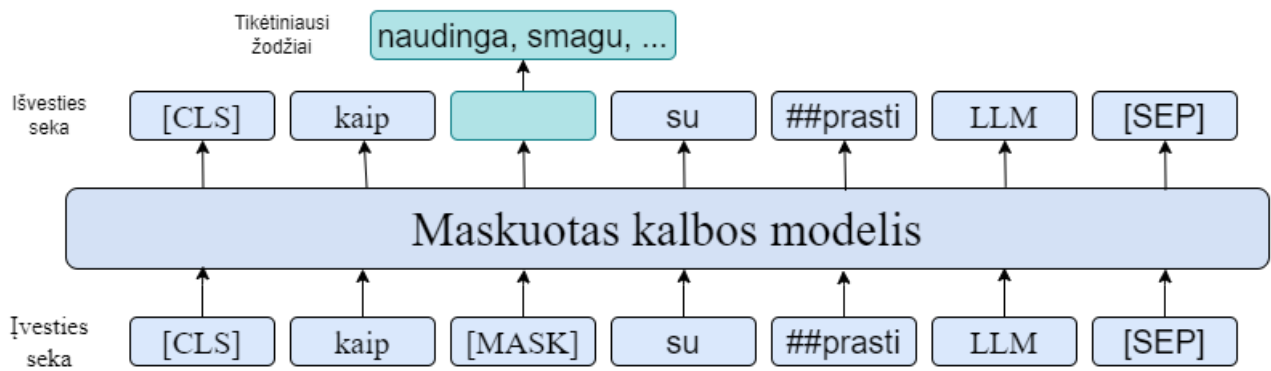
$$\text{įvertis} = \frac{\text{dažnis}_{\text{pora}}}{\text{dažnis}_{1\text{elementas}} \cdot \text{dažnis}_{2\text{elementas}}} \quad (8)$$

Palyginus poros dažnį su kiekvienos jos dalių dažnių sandauga, algoritmas teikia pirmenybę porų jungimui, kurių individualios dalys yra rečiau pasitaikančios žodyne. Toks skaičiavimas padeda užfiksuoti morfologinę informaciją ir apibendrinti žodžius tarp skirtingų formų. Algoritmas nerasdams tam tikro žodžio savo žodyne – jį pažymi '*unknown*' žyma, tačiau prieš tai bando aptikti galimai turimas žodžio dalis. Bendras žodyno sudarymo procesas yra gana sudėtingas ir reikalauja didelių skaičiavimo resursų, nes kiekvienoje iteracijoje yra pasirenkama optimaliausia pora [43]. Nepaisant to, dėl puikių rezultatų, gebėjimo apdoroti nežinomus žodžius bei prisitaikymo prie sudėtingų kalbų, jis yra gana populiarus žetonizavimo algoritmas taikomas transformerių modelių architektūrose.

1.5.1.3. Išankstinis mokymas

BERT modelis išankstinio apmokymo metu yra apmokomas taikant du neprižiūrimo mokymo tipo uždavinius. Prieš pristatant BERT modelį, transformerių tipo modeliai buvo orientuoti į sekančio žodžio prognozavimą, todėl jie buvo apmokinami apdorojant tekstą nuosekliai iš kairės į dešinę. BERT architektūroje pristatytas naujas metodas, pavadintas maskuotu kalbos modeliu (angl. *masked language model*, MLM) [44]. Antrasis mokymosi etapas, , buvo orientuotas į sekančio sakinio prognozavimą siekiant sukurti modelį, gebantį atpažinti ryšius tarp sakinių.

MLM atsitiktinai užmaskuoja žodžius sakinyje, o tada bando juos atspėti. Maskavimas reiškia, kad modelis bandydamas prognozuoti užslėptą žodį, vertina visą sakinio kontekstą, ir atsižvelgia į žodžius iš kairės ir dešinės [39].



1.5.2 pav. Maskuoto kalbos modelio įvesties ir išvesties sekų atvaizdavimas (sudaryta remiantis [45])

BERT modelio apmokymo metu atsitiktinai pasirenkama ir užmaskuojama 15% visų *WordPiece* sugeneruotų žetonų. Tačiau dėl pasirinktos maskavimo technikos atsiranda neatitikimai tarp modelio apmokymo ir parametrų redagavimo etapų. Po modelio apmokymo *[MASK]* žetonas nėra naudojamas pritaikymo ir koregavimo etapuose. Dėl to, ne visi atrinkti maskuojami žodžiai yra pakeičiami *[MASK]* žetonu [39]. Generuojant mokymo duomenis, 80% atvejų pasirinkta žyma pakeičiama *[MASK]* reikšme, 10% atvejų keičiama atsitiktine žyma ir 10% atvejų pasirinkta žyma lieka nepakeista.

Antroji BERT apmokymo dalis skirta sugebėti prognozuoti sekantį sakinį tekste. Pirmajame etape atliekamas kalbos modeliavimo apmokymas suteikia kalbos suvokimo pagrindus, tačiau šių žinių modeliui siekiant suprasti sakinių ryšius. Sakinių susiejimo supratimas yra svarbi daugelio NLP uždavinių sprendimo dalis. NSP užduoties apmokymui buvo pasirinkta plati duomenų imtis, kurioje sakiniai suskirstyti į *A* ir *B* kategorijas. Apmokymo metu buvo pasirinkta 50% *B* sakinių, kurie tikslingai sekė *A* sakinį bei 50% atsitiktinių *B* sakinių, kurie buvo suklasifikuoti kaip ne sekantys sakiniai [39]. Šis mokymosi etapas leidžia modeliui įgyti geresnį supratimą apie teksto struktūrą ir kontekstą.

Visa pagrindinio BERT modelio apmokymo duomenys surinkti iš *BookCorpus* ir anglų kalbos *Wikipedia* internetinių svetainių. Šie šaltiniai buvo pasirinkti didelės turinio apimties ir plataus konteksto. Vėliau buvo sukurtos kitos BERT modelio atšakos, kurios buvo apmokytos naudojant įvairiakalbius žodynus siekiant išplėsti modelio panaudojimo galimybes.

1.5.1.4. Modelio koregavimas

BERT modelio koregavimas ir pritaikymas konkrečių uždavinių sprendimui yra paprastesnis ir mažiau išteklių reikalaujantis procesas, palyginti su apmokymo etapu. Koreguojant BERT modelį tam, kad jis galėtų spręsti klasifikavimo užduotis, pavyzdžiui, sentimentų analizę, į modelio architektūrą įtraukiamas papildomas klasifikavimo sluoksnis. Šio etapo metu naujas sluoksnis yra vienintelis sluoksnis, kuriam parametrai yra inicializuojami naujai. Visi kiti modelio parametrai yra perimami iš išankstinio apmokymo etapo.

Klasifikavimo užduotyse siekiant atvaizduoti visą įvesties kontekstą BERT modeliuose naudojama pirmojo *[CLS]* žetono iš paskutinio sluoksnio *h* būseną. Tada pritaikoma nesudėtinga eksponentinio normalizavimo funkcija (žr. (9) formulė), skirta nustatyti atitinkamos kategorijos priskyrimo tikimybę [46].

$$p(c|h) = \text{softmax}(W \cdot h); \quad (9)$$

čia W – specifinės užduoties parametų matrica.

Koregavimo metu yra modifikuojami visi BERT modelio parametrai. Tačiau pirmieji 7-8 modelio sluoksniai nedaro reikšmingos įtakos galutiniams klasifikavimo rezultatams. Šie sluoksniai yra linkę išmokti ir suprasti pagrindines kalbos dalis, kurios turi mažiau įtakos specifinių užduočių rezultatuose [47]. Taikant modelio parametų koregavimą, svarbu atsižvelgti į keletą BERT modelio ypatybių [46]:

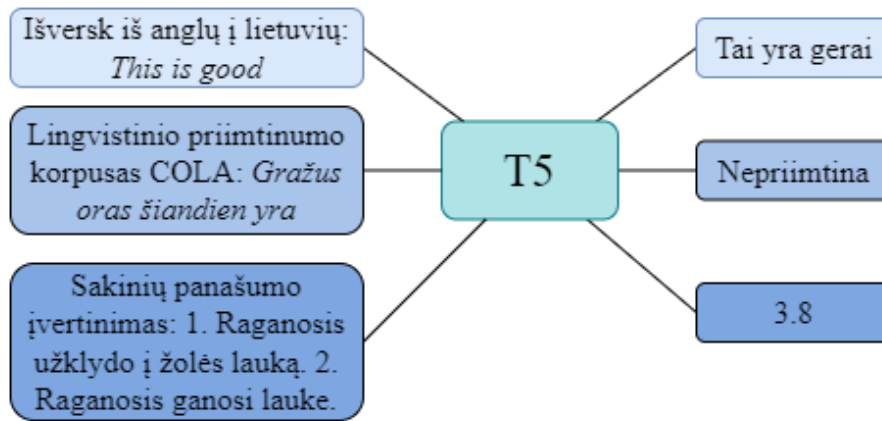
- ribotas palaikomų žetonų konteksto ilgis yra 512;
- reikalinga tinkamai identifikuoti klasifikavimui parenkamą sluoksnį ir jo tipą, BERT_{BASE} modelis sudarytas iš žodžių įterpimo, dvylikos koduotuvų ir sutelkimo sluoksnio;
- susiduriama su persimokymo rizika parametų koregavimo metu, dėl nedidelės duomenų imties ar netinkamai pasirinkto optimizavimo algoritmo.

Naudojant BERT_{LARGE} modelį ar jo modifikacijas, maži duomenų rinkiniai gali grąžinti nestabilius rezultatus. Pabrėžtina, kad įvairios BERT modelio modifikacijos, pavyzdžiui, *UNCASED BERT*, *ROBERTA_{LARGE}* ir kitos, taip pat yra jautrios duomenų rinkinio pasiskirstymui apmokymo metu [48]. Rekomenduotina išbandyti įvairiai paskirstytas duomenų imtis, tačiau tai kartais sukelia kitas rizikas, kaip per didelis prisitaikymas prie duomenų.

Nepaisant iššūkių, didelių kalbos modelių, tokių kaip BERT, koregavimas laikomas vienu efektyviausiu būdu taikyti giliojo mašininio mokymosi modelių algoritmus NLP uždaviniams spręsti. Šis metodas leidžia sutaupyti daug resursų, greičiau gauti rezultatus bei bendrai pasiekti geresnius rezultatus. BERT modelis, naudojamas kaip pritaikymo pagrindas, leidžia efektyviai apdoroti kalbos kontekstą. Papildomų sluoksnių panaudojimas BERT modelio architektūroje suteikia galimybę modelį pritaikyti įvairioms NLP užduotims, kaip teksto klasifikavimas, vardinių objektų atpažinimas ar klausimų atsakymas.

1.5.2. T5 tipo modeliai

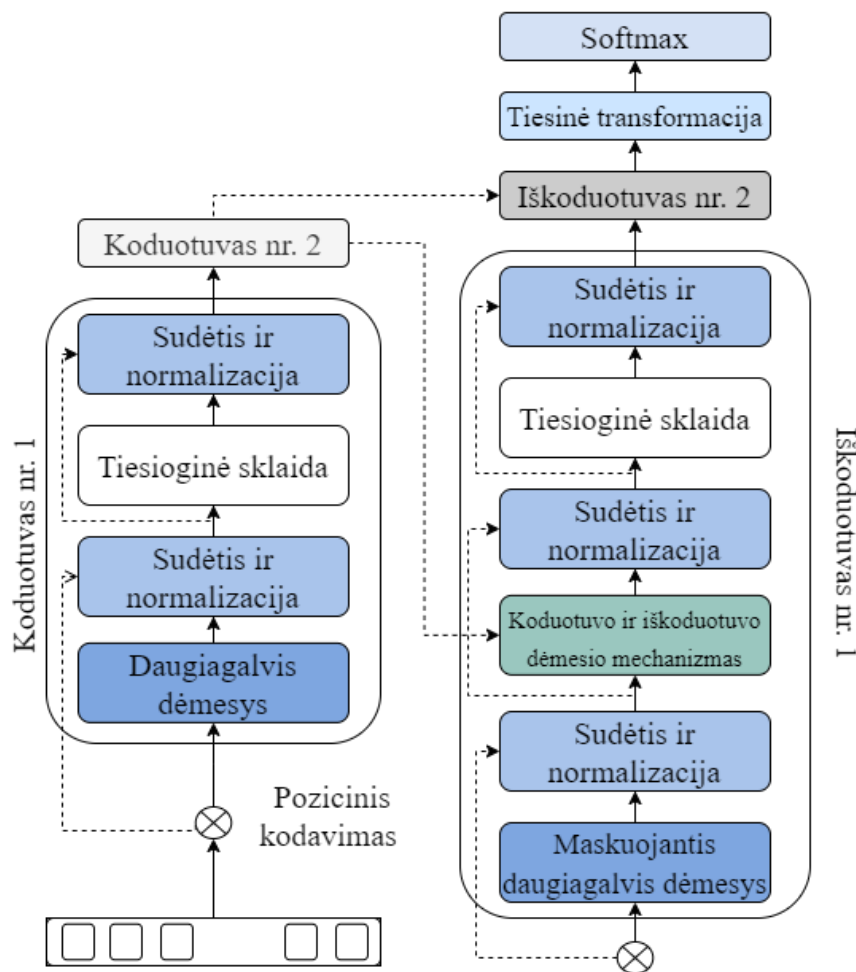
Iš teksto į tekstą perdavimo transformatorius (angl. *text-to-text transfer transformer*, T5) – tai transformerio modelio architektūra skirta įvairiems natūralios kalbos apdorojimo uždaviniams spręsti [29]. Šis modelis buvo pristatytas 2019 metais. Šis modelis yra išskirtinis tuo, kad sprendžia visus jam pateiktus uždavinius remdamasis principu *„iš teksto į tekstą“*. 1.5.3 pav. diagramoje pateikiamas T5 modelio veikimo apibendrinimas. Visi nagrinėjami uždaviniai, įskaitant vertimą, klasifikavimą bei klausimų atsakymą, yra pritaikomi teksto formulavimo uždaviniui. Tai reiškia, kad nepaisant užduoties pobūdžio, modeliui tiek įvesties duomenys, tiek norimi gauti rezultatai yra pateikiami tekstiniu formatu. Toks uždavinių sprendimo būdas suteikia galimybę naudoti vieną modelį, vienodą nuostolių funkciją ir hiperparametrus, visuose užduočių aibėse [29].



1.5.3 pav. T5 modelio veikimo apibendrinimas (sudaryta remiantis [29])

1.5.2.1. Modelio struktūra

T5 modelio architektūra yra paremta tradicine koduotuvo–iškoduotuvo transformerio architektūra, kurios vieno bloko struktūra iliustruota 1.5.4 pav.



1.5.4 pav. T5 modelio architektūra (sudaryta remiantis [49])

Pagrindinis sistemos modelis buvo sukurtas alternatyvaus dydžio BERT_{BASE} modelio architektūrai. 1.5.4 pav. matoma, kad T5 modelio architektūroje kiekvienas blokas turi savo dėmesio mechanizmą, koduotuvo ar iškoduotuvo dėmesio sutelkimo mechanizmą ir tiesioginio sklaidimo neuroninį tinklą.

Kiekviename bloke neuroninio tinklo modelis susideda iš tankaus sluoksnio su išvesties dimensijos dydžiu d_{ff} , kurį seka netiesinė ReLU projekcija ir dar vienas tankus sluoksnis [29].

T5 modelis buvo pristatytas su penkiomis skirtingomis modelio architektūros versijomis: T5_{SMALL}, T5_{BASE}, T5_{LARGE}, T5_{3B}, T5_{11B}. Pagrindiniai parametrai, nusakantys modelio dydį ir parametrų skaičių, yra [29]:

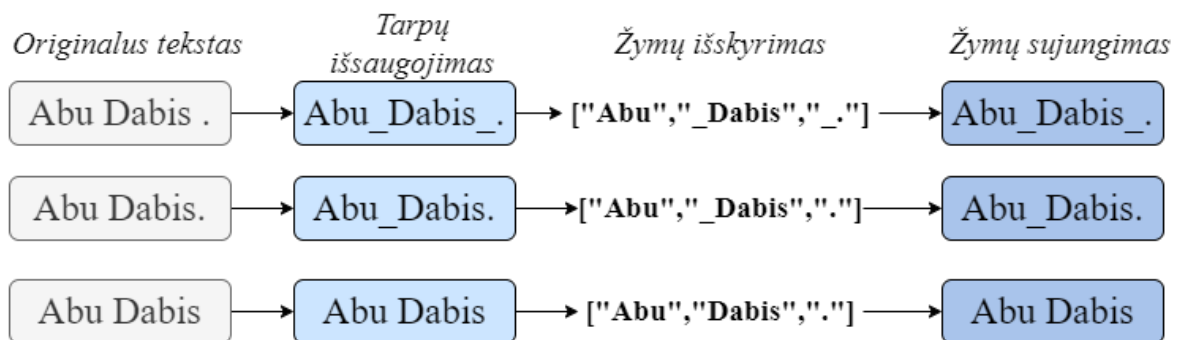
- n_{blocks} – transformerio blokų skaičius;
- d_{ff} – tarpinio vektoriaus dimensijų skaičius neuroniniame tinkle;
- d_m – įterpimo vektoriaus dimensijų skaičius, kuris atitinka transformerio bloko išvesties vektoriaus dydį;
- d_{kv} – *Key* ir *Value* matricų, kurios naudojamos dėmesio sutelkimo mechanizme, dimensijų skaičius;
- n_h – dėmesio mechanizmuose naudojamų dėmesio sutelkimo galvų skaičius.

T5_{BASE} modelį sudaro 220 milijonų parametrų, kur $n_{blocks} = 12$, $d_{ff} = 3072$, $d_m = 768$, $d_{kv} = 64$, $n_h = 12$, o T5_{LARGE} modelis yra sudarytas iš 738 milijonų parametrų, kur $n_{blocks} = 24$, $d_{ff} = 4096$, $d_m = 1024$, $d_{kv} = 64$, $n_h = 16$.

1.5.2.2. Žetonizacija

T5 modelis žetonų išskyrimui naudoja *SentencePiece* algoritimą. Tai paprastas, efektyvus ir nepriklausomas nuo kalbos žodžių žymų išskyrimo metodas. T5 modelyje naudojamą *SentencePiece* žodyną sudaro 32,000 žetonų. Išskirtinės taikomos savybės algoritmui [50]:

- informacijos neprarandantis žetonų išskyrimas;
- išskiriami ir išsaugomi specialūs simboliai, kuriais nusakomi nežinomi žodžiai *<unk>*, sekos pradžia *<s>* ir pabaiga *</s>*;
- daugiakalbio apmokymo metu galima apibrėžti papildomus simbolius identifikuojančius specifinę kalbą;
- simbolių normalizavimui taikomas *Unicode NFKC* metodas.



1.5.5 pav. Informacijos neprarandančios žetonizacijos algoritmo pavyzdys (sudaryta remiantis [41])

SentencePiece metodas žetonų išskyrimo metu išsaugo visą reikiamą informaciją norint atkurti pradinę duomenų įvestį. Pagrindinis informacijos neprarandančios žetonizacijos algoritmo principas yra vykdomas, laikant įvesties tekstą unikodo simbolių seka. Visi simboliai, įskaitant skyrybos ženklus, yra vertinami kaip vienodos reikšmės simboliai. *SentencePiece* pirmiausia pakeičia tarpą

meta simboliu ‘_’ ($U+2581$), o vėliau atlieka žetonų išskyrimo procesą [50]. Norint atkurti originalų tekstą, telieka meta simbolį ‘_’ pakeisti tarpu.

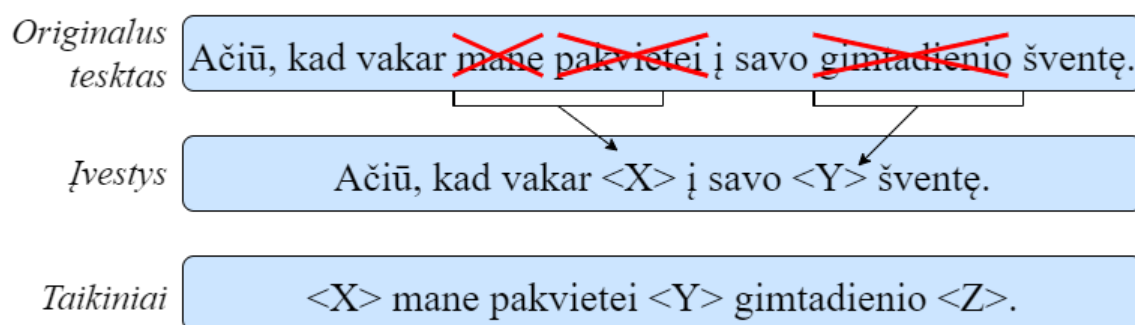
SentencePiece pritaiko greitaveikos procesus apmokymui ir segmentacijai. Duotam įvesties sakiniui ilgio N , algoritmas pareikalauja tik $O(N \log(N))$ skaičiavimo sąnaudų. Šis algoritmas simbolių sujungimui naudoja prioritetų eilę [50]. Dėl didelės greitaveikos, naudojant *SentencePiece* nereikalinga išankstinė specifinės kalbos žetonizacija. Gebėjimas greitai ir efektyviai išskirti ir sudaryti žetonų daugiakalbį žodyną paverčia šį algoritmą labai patogiu naudoti daugiakalbiuose modeliuose.

1.5.2.3. Išankstinis apmokymas

T5 modelis sukurtas remiantis visų problemų sprendimu taikant ‘iš teksto į tekstą’ metodiką. Tai leidžia modelį projektuoti taip, kad prognozavimo metu būtų taikoma įprastinė didžiausio tikėtimumo metodika, taikant priverstinio mokytojo mokymosi metodiką (angl. *teacher forcing*) ir skaičiuojamas kryžminės entropijos nuostolis. Apmokymo metu optimizavimui naudojamas *AdaFactor* algoritmas, o testavimo etape taikomas godaus pobūdžio atsakymų dekodavimas [29].

Išankstinio apmokymo metu, pirmajam T5 modelio apmokymui buvo naudotas internetine surinktas tekstas, pasiekiamas iš atvirojo kodo resurso *Common Crawl* svetainėje. Prieš apmokant T5 modelį, buvo atliktas minimalus duomenų išvalymas, pašalinant eilutės, kurios [51]:

- nesibaigė sakinio pabaigos ženklu;
- turi žodį *JavaScript*;
- yra tarp ‘{’ bei ‘}’ ženklų.



1.5.6 pav. T5 išankstinio mokymosi uždavinio pavyzdys (sudaryta remiantis [29])

T5 modelis išankstinio apmokymo etape yra mokomas suprasti bendrines kalbos žinias. Apmokymo procesas yra paremtas MLM mokymu. T5 apmokymo etapas yra sukurtas atsižvelgiant į BERT taikomus metodus bei žetonų išmetimo technikas. Apmokymo etape duomenų paruošimas vykdomas šia veiksmų seka (žr. 1.5.6 pav.) [29]:

- atsitiktinai išrenkama 15% įvesties sekos žetonų;
- visi iš eilės einantys žetonai yra pakeičiami vienu saugos (angl. *sentinel*) ženklu. Kiekvienam saugos ženklui priskiriama unikali reikšmė įvesties sekos ribose. Šie žetonai yra specialaus pobūdžio žymos, atpažįstamos modelio žodyno rinkinyje;
- sukuriamas tekstinis uždavinio taikinytis, kuris sudaromas sujungiant visus įvesties žetonus atskirtas saugos žymomis;
- sudarytas galutinio taikinio tekstas yra baigiamas specialiuoju pabaigos žetonu <Z>.

T5 modelių iš ankstinis apmokymas atliekamas per $2^{19} = 524,288$ žingsnių. Maksimalų žetonų sekos ilgis apribojamas iki 512. Apmokymo rinkinį sudaro 128 sekos, siekiant, kad rinkinys apytiksliai sudarytų $2^{16} = 65,536$ žetonus. Iš viso išankstinis T5_{BASE} modelio apmokymas buvo atliktas su ≈ 34 milijardais žetonų.

1.5.2.4. Modelio koregavimas

T5 modelio koregavimas bei pritaikymas NLP užduotims remiasi ta pačia strategija, kuri buvo taikoma modelio išankstinio apmokymo metu. Kiekvienai užduočiai, kuriai yra pritaikomas T5 modelis, yra atliekamas tolimesnis individualiai pritaikytas apmokymas prieš įvertinant modelio efektyvumą. Pagrindinis modelio koregavimo etapo žingsnis yra individualios užduoties paruošimas į *„iš teksto į tekstą“* formatą. Tai reiškia, kad tikėtini atsakymai, turi būti pateikti kaip tekstinė seka [51]. Dėl šios priežasties, vienas modelis gali būti mokomas prognozuoti skirtingų uždavinių rezultatus.

Priešingai nei BERT modelyje, atliekant T5 modelio koregavimą, modelio architektūroje papildomi pokyčiai nėra atliekami. Priklausomai nuo atitinkamų parametrų, kaip siekiamas rezultatų tikslumas, stabilumas, duomenų imties dydis ar invariantiškumas, mokymosi metu galima atitinkamai taikyti dalinį arba etapais grystą sluoksnių parametrų kitimo (angl. *gradual unfreezing*) metodiką [29]. Ši metodika leidžia koreguoti vis daugiau modelio parametrų, keičiantis mokymosi epochoms. Tai gali padėti pagreitinti modelio pritaikymą specifinėms užduotims. Tačiau tai tik vienas iš galimų modelio koregavimo būdų, naudojant T5 modelį galima taikyti įvairias reguliavimo strategijas.

1.6. Rezultatų vertinimo metrikos

Tolimesniuose skyriuose detalizuojamas sentimentų analizės modelio tyrimas, pateikiami rezultatų palyginimai ir aptariami modelių įvertinimo rezultatai, remiantis toliau aprašomomis metrikomis:

1. *kategorinis kryžminės entropijos nuostolis* – naudojamas modelių mokymosi metu siekiant įvertinti modelio prognozių tikslumą bei atitinkamai reguliuoti modelio svorius [52]. Ši metrika dažnai naudojama klasifikavimo problemose siekiant įvertinti modelio prognozių pasiskirstymo atitikimą tikrajam duomenų pasiskirstymui [24]. Mažėjanti kryžminės entropijos reikšmė rodo, kad modelis mokymosi metu išmoksta geriau atspindėti duomenis. Kryžminės entropija $H(p, q)$ tarp dviejų tikimybių pasiskirstymų p ir q apskaičiuojama pagal (10) formulę [24].

$$H(p, q) = - \sum_x p(x) \cdot \log(q(x)); \quad (10)$$

čia $p(x)$ yra tikrasis tikimybės pasiskirstymas, o $q(x)$ yra modelio numatytas tikimybės pasiskirstymas. Mokymosi metu siekiama sumažinti šį nuostolio įvertį: kuo mažesnis nuostolis, tuo geriau modelis geba prognozuoti. Ši metrika yra viena iš dažniausiai naudojamų metrikų neuroninių tinklų modelių įvertinimui, kai taikoma *softmax* funkcija [52]. Ji puikiai pritaikoma daugiaklasių klasifikavimo uždaviniams stebėti.

2. *tikslumas* – tai rodiklis, kuris nesudėtingai įvertina modelio gebėjimą prognozuoti kategorijas. Ši metrika nusako santykį tarp modelio prognozuojamų kategorijų ir tikrųjų įrašų reikšmių. Ji ypač naudinga, kai visos galimos klasės turi vienodą pasiskirstymą ir yra vienodai svarbios. Ši metrika apskaičiuojama remiantis (11) formule [53].

$$Tikslumas = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

Formulėje matomos keturios skirtingos klasifikavimo reikšmės. TP – tai teisingo priėmimo (angl. *true positive*) įvertis, kuris nusako, kad priskirta klasė buvo teisingai identifikuota. TN – tai teisingo atmetimo (angl. *true negative*) įvertis, kuris nusako, kad modelis teisingai nepriskyrė atitinkamos klasės. O klaidingo priėmimo FP (angl. *false positive*) bei klaidingo atmetimo FN (angl. *false negative*) reikšmės nusako, kad modelis priskyrė netinkamas kategorijas atitinkamoms įvestims [53]. Todėl apibendrintai, ši tikslumo metrika nurodo, kokia dalis visų gautų modelio prognozuojamų kategorijų buvo teisingai suklasifikuotos. Ši metrika yra nesudėtinga vertinti bei palyginti, tačiau ji neatsižvelgia į netolygų klasių pasiskirstymą duomenų imtyje.

3. $F1$ įvertis – tai alternatyvi mašininio mokymosi vertinimo metrika, kuri įvertina modelio gebėjimą kategorizuoti įvestis atsižvelgiant į skirtingų klasių pasiskirstymą [54]. Rezultatų analizė pagal atskiras klases pagerina įvertinimo kokybę lyginant su anksčiau aprašyta tikslumo metrika. $F1$ įvertis apjungia du vertinimo rodiklius: preciziškumą (angl. *precision*) ir atkūrimą (angl. *recall*) [54]. Modelio įverčio metrikos gali įgyti reikšmes tarp 0 ir 1, kur 1 geriausias įmanomas įvertis. Remiantis kryžminės matricos įvertinimo metrikomis TP , TN , FP bei FN , $F1$ įvertis apskaičiuojamas remiantis (14) formule [55].

$$Preciziškumas = \frac{TP}{FP + TP} \quad (12)$$

$$Atkūrimas = \frac{TP}{FN + TP} \quad (13)$$

$$F_1 = 2 * \frac{Preciziškumas \cdot Atkūrimas}{Preciziškumas + Atkūrimas}; \quad (14)$$

čia $F1$ įvertis priklauso nuo (12) formulėje aprašyto preciziškumo ir (13) formulėje aprašyto atkūrimo rodiklių. Šie įverčiai priklausomi nuo TP , TN , FP bei FN įverčių per visas klases. Šios metrikos leidžia apskaičiuoti bendrą $F1$ modelio įvertį [55]. Toliau sentimentų analizės modeliui taikomas pavyzdžių svoriu paremtas $F1$ įvertis (angl. *sample-weighted F1 score*), kuris yra pritaikomas vertinant prognozes, kur duomenys pagal kategorijas yra pasiskirstę netolygiai [54]. Pavyzdžių svorių $F1$ įvertis apskaičiuojamas, kaip svorio vidurkis pagal atskirų klasių $F1$ įverčius.

2. Sistemos projektas

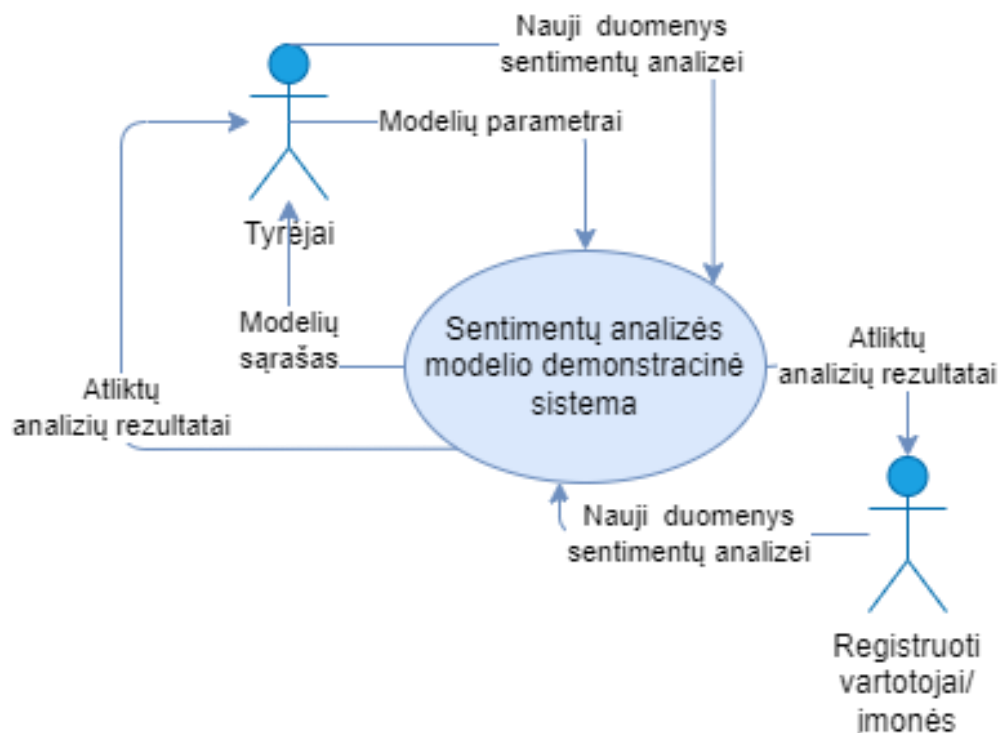
Šiame skyriuje pateikiami pagrindiniai programinės įrangos techniniai aspektai, įgyvendinti magistratūros studijų metu. Magistro studijų metu buvo orientuojamasi į modelio, paremto giliuoju mašiniu mokymusi, pritaikymą lietuviškų komentarų sentimentų analizei. Iki šiol lietuvių kalbos sentimentų analizės tema yra atlikta tik keletas tyrimų. Siekiant analizuoti modelio funkcionalumą, gebėjimą juo naudotis, palyginti eksperimentinius modelius ir stebėti rezultatų kitimą laike, darbo metu buvo suprojektuota ir įgyvendinta demonstracinės paskirties informacinė sistema. Ši prototipinė sistema yra skirta sentimentų analizės modelių testavimui ir pritaikyta potencialiems klientams bei susidomėjusiems mokslininkams. Toliau detalizuojami sistemos reikalavimai, aprašomi pagrindiniai architektūriniai sprendimai bei sukurto prototipo rezultatai.

2.1. Sistemos veiklos sudėtis

Šiame skyriuje aprašomas sukurtos sistemos veiklos kontekstas, jo veiklos įvykių detalizavimas ir sistemos panaudojimo atvejai.

2.1.1. Sistemos kontekstas

Bendrais bruožais sistemos konteksto diagrama pateikiama 2.1.1 pav.:



2.1.1 pav. Prototipinės sentimentų analizės sistemos konteksto diagrama

Konteksto diagramoje yra išskiriami 2 pagrindiniai vartotojų tipai, kurie dalinasi skirtingų tipų duomenų srautais reikalingais sentimentų analizės atlikimui. Diagramoje pateikiama visų sistemos vartotojų galimybė atlikti norimų duomenų sentimentų analizę. Tokiu būdu išbandomas modelio veikimas bei prognozavimo galimybės. Svarbiausia sistemos dalis yra laisvo formato teksto duomenų analizavimas ir rezultatų pateikimas siekiant kuo tikslesnio sentimentų identifikavimo. Veiklos konteksto diagramoje nėra teikiamas dėmesys gretutiniam sistemos funkcionalumo duomenų srauto vaizdavimui, kaip vartotojo duomenys, nustatymai ir kita. Kuriama sistema veiks kaip atskira sistema, kuri galės pateikti modelio veikimą be papildomo integravimo su išorinėmis sistemomis.

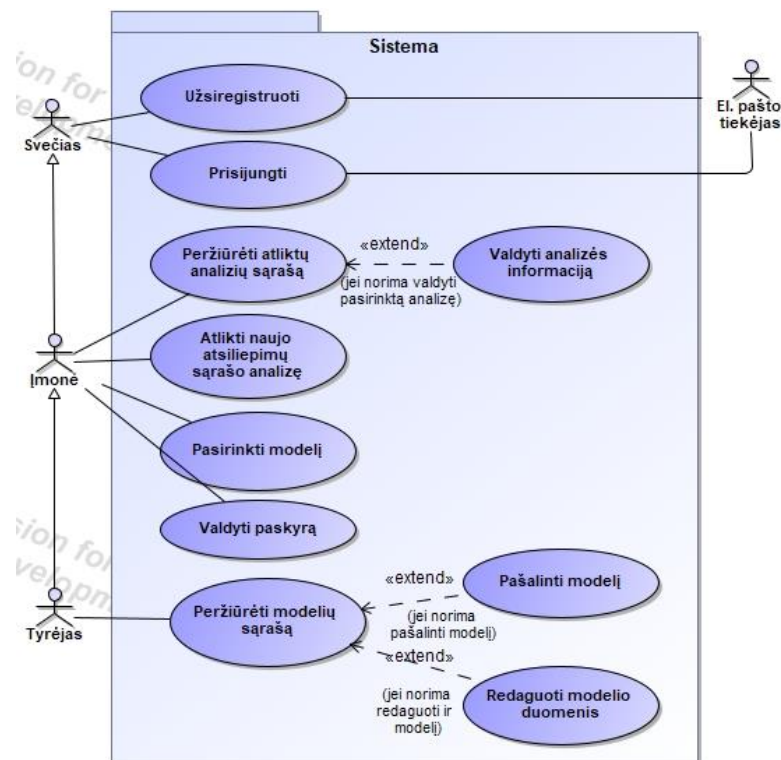
2.1.2. Sistemos veiklų padalijimas

Detalizuoti veiklos srities duomenų paaiškinimai pateikiami 2.1.1 lentelėje:

2.1.1 lentelė. Sistemos veiklos įvykių sąrašo detalizavimas

Nr.	Srauto pavadinimas	Srauto kryptis		Aprašymas
		Iš	Į	
1.	Nauji duomenys sentimentų analizei	Tyrėjas	Demonstracinė sentimentų sistema	Tyrėjai sistemoje turės galimybę įkelti naujus lietuviškų atsiliepimų duomenis.
2.	Modelių parametrai	Tyrėjas	Demonstracinė sentimentų sistema	Tyrėjai sistemoje turės galimybę valdyti tam tikrus sentimentų modelio parametrus.
3.	Modelių sąrašas	Demonstracinė sentimentų sistema	Tyrėjas	Prireikus, tyrėjai sistemoje galės peržiūrėti visų sistemoje registruotų modelių sąrašus ir jų parametrus.
4.	Atliktų analizių rezultatai	Demonstracinė sentimentų sistema	Tyrėjas	Tyrėjai sistemoje turės galimybę matyti atliktų analizių rezultatus.
5.	Atliktų analizių rezultatai	Demonstracinė sentimentų sistema	Įmonės/registruoti vartotojai	Registruoti įprasti vartotojai sistemoje turės galimybę įkelti naujus lietuviškų atsiliepimų duomenis.
6.	Nauji duomenys sentimentų analizei	Įmonės/registruoti vartotojai	Demonstracinė sentimentų sistema	Registruoti įprasti vartotojai sistemoje turės galimybę matyti atliktų analizių rezultatus.

2.1.3. Sistemos panaudojimo atvejai



2.1.2 pav. Prototipinės sentimentų analizės sistemos panaudojimo atvejų diagrama

Bendraisiais bruožais sistemos veikimas yra pateikiamas jos numatomų panaudojimo atvejų diagrama (žr. 2.1.2 pav.). Diagramoje pateikiama skirtingų vartotojų sąveika su sistema. Šioje

diagramoje yra trys vartotojų tipai: *svečias*, *įmonė* ir *tyrėjas*, bei *elektroninio pašto tiekėjas* kaip išorinė sistema. Visi sistemos naudotojai paveldi *svečio* panaudojimo atvejus:

- užsiregistruoti;
- prisijungti.

Įmonės tipo vartotojai gali išbandyti įvairius sistemoje pateikiamus sentimentų analizės modelius. Įmonių panaudojimo atvejai:

- peržiūrėti atliktų analizių sąrašą;
- valdyti analizės informaciją;
- atlikti naują atsiliiepimų sąrašo analizę;
- pasirinkti modelį.

Tyrėjai sistemoje turėdami didesnes valdymo teises, gali analizuoti sistemos sentimentų analizės modelius, konfigūruoti jų parametrus, valdyti modelio versijas ir saugyklos nustatymus. Papildomi *tyrėjo* tipo vartotojų panaudojimo atvejai:

- peržiūrėti modelių sąrašą;
- pašalinti modelį;
- redaguoti modelio parametrus, konfigūraciją.

2.2. Reikalavimai sistemos funkcionalumui

Šiame skyriuje aprašomi esminiai sistemos funkciniai ir nefunkciniai reikalavimai, skirti užtikrinti, kad sistema atitiktų vartotojų lūkesčius bei veiktų efektyviai ir patikimai.

2.2.1. Funkciniai reikalavimai

Remiantis sistemos kūrimo tikslais bei siekiant užtikrinti tenkinantį sistemos naudojimo funkcionalumą, sistemai apibrėžti detalūs funkciniai reikalavimai. Toliau pateikiami tik esminiai sentimentų analizės atlikimui reikalingi funkciniai reikalavimai remiantis projekto techninės dokumentacijos aprašymu:

- sistema privalo suteikti galimybę vartotojams išsaugoti savo atliktus darbus bei juos peržiūrėti ir analizuoti kitu laiku (funkcinis reikalavimas nr. 3);
- sistema privalo suteikti galimybę įkelti lietuviškų atsiliiepimų tekstus vartotojams, pasirinktus duomenis analizuoti naudojantis sistemos modeliais ir analizės rezultatus tvarkyti pagal savo poreikius (funkcinis reikalavimas nr. 4 ir nr. 5);

2.2.2. Nefunkciniai reikalavimai

Detalūs sistemos nefunkciniai reikalavimai yra apibrėžti projekto techninėje dokumentacijoje, kuri buvo pristatyta magistro studijų metu. Toliau šiame skyriuje pateikiami esminiai nefunkciniai reikalavimai projektuojamai sistemai.

2.2.2.1. Reikalavimai sistemos išvaizdai

Kuriama prototipinio pobūdžio sistema, kuri turi būti sukurta remiantis šiais stiliaus kriterijais:

- *profesionali sistemos išvaizda*. Kuriama oficiali tinklalapio struktūra. Komponentų išdėstymas ir atvaizdavimas turi būti aiškiai pabrėžiamas ir nenaudojami pertekliniai judantys

komponentai sąsajai pagyvinti. Naudojimosi mygtukai turi būti aiškiai matomi, siekiama, kad vartotojas turėtų gebėti naviguoti neturėdamas didelės patirties;

- *sąsaja aktyviai bendraujanti su vartotoju*. Dirbdamas su sistema, vartotojas turi matyti savo progreso rezultatus. Sistema turi informuoti apie sėkmingai atliktas operacijas, o nesėkmės atveju – turi būti pateikiami pakankamai informatyvūs klaidų pranešimai..

2.2.2.2. Reikalavimai vykdymo charakteristikoms

Toliau pateikiami esminiai nefunkciniai reikalavimai sistemos vykdymo veiklų procesams:

- *sentimentų analizės įrašo rezultatai sugeneruojami per greičiau nei 10s*. Norint atlikti įvairius bandymus tikimasi, kad sistemos greitaveika neapribos tyrimų darbo našumo. Laiko apdorojimo trukmė apibrėžiama tekstams atitinkantiems modelio bei sistemos duomenų apribojimo reikalavimus;
- *sistema turi tinkamai apdoroti .csv, .txt bei .json formato failus*. Sistemoje vartotojai duomenis analizei privalo galėti pateikti apibrėžtu formatu, tačiau įvairaus tipo duomenų failais. Apdoroti ir išanalizuoti duomenys sistemoje saugomi bei vartotojui pateikiami .csv formatu.

2.3. Sistemos vystymo procesas

Sistemos kūrimo procesas išskiriamas į tris pagrindinius žingsnius: specifیکavimą, projektavimą ir realizavimą. Visuose projekto įgyvendinimo etapuose darbas periodiškai peržiūrimas, vertinamas ir atitinkamai atsižvelgiama į galimus pasikeitimus dėl naujų žinių, reikalavimų suvokimo bei laiko ribotumo.

Sistemos įgyvendinimo etapas vykdomas remiantis modifikuota *Agile* metodologija. Metodologija pritaikoma supaprastintam iteraciniam programinės įrangos kūrimo procesui. Numatoma, kad kiekviena iteraciją sudarys: darbo planavimas, įgyvendinimas, funkcionalumo testavimas, demonstracijos pristatymas. Bendruoju atveju numatomas realizavimo procesas apima sistemos kūrimą vienodos trukmės iteracijomis bei periodišku darbo pristatymu vadovui bei kitiems suinteresuotiems asmenims.

2.4. Sistemos architektūros specifیکacija

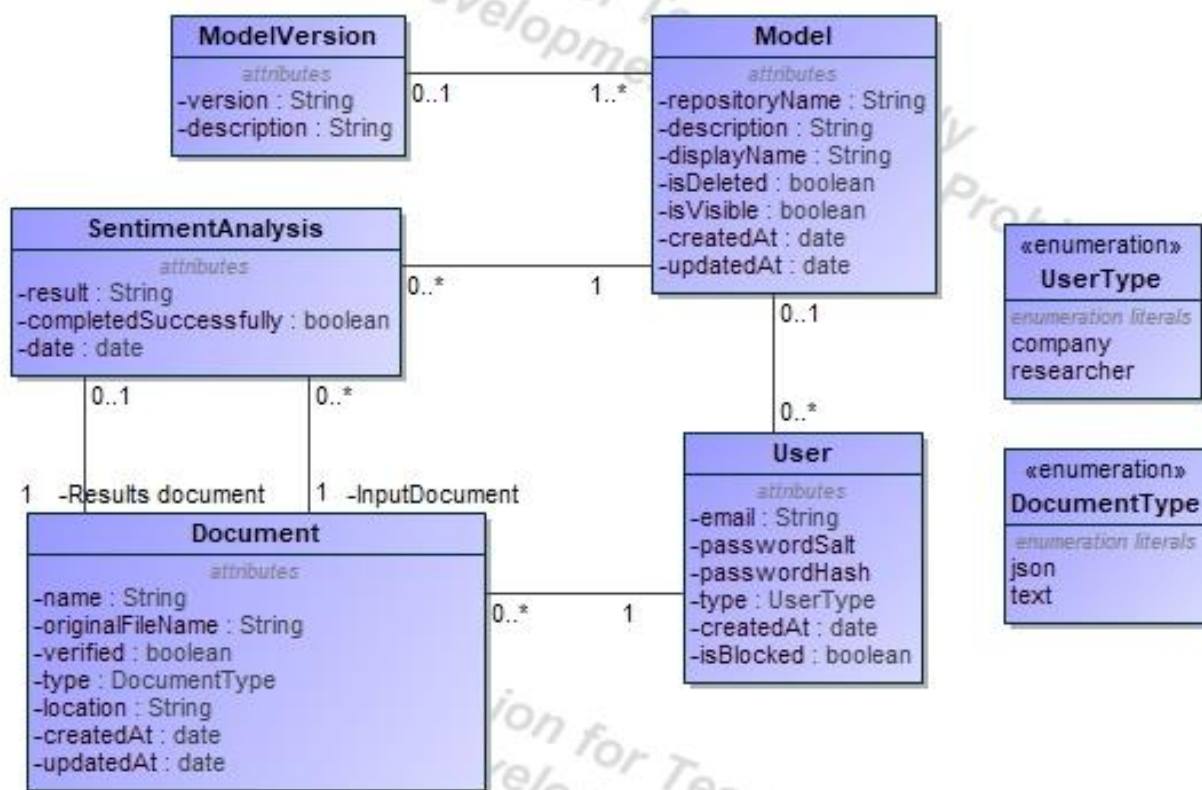
Šiame skyriuje yra pateikiama apibendrinta įgyvendintos programinės sistemos architektūros apžvalga, apibrėžiant pagrindinius komponentus, sąveikas ir projektavimo sprendimus. Dokumente pateikiami prototipinės modelio sistemos komponentai, sąveikos tarp posistemų, pagrindiniai duomenų modeliai ir srautai.

2.4.1. Esybių ryšių modelis

Dalykinės srities objektų specifیکacija yra atliekama projektuojant bei analizuojant esybių ryšių modelį. 2.4.1 pav. yra pateikiamas projektuojamos sistemos esybių ryšių modelis projektuojamai sistemai. Pagrindinės suprojektuotos sistemos esybės:

- *Model* – tai giliuoju mokymusi paremtų sentimentų analizės modelių duomenų saugojimo esybė, kurioje laikoma informacija apie modelį, jo saugyklos duomenis, versijas;
- *Document* – tai esybė, kuri yra skirta saugoti informaciją apie vartotojų įkeltus dokumentus ir atvirojo tipo tekstus;

- *User* – tai bendrojo pobūdžio esybė, kuri yra skirta saugoti informaciją apie sistemoje užsiregistravusius vartotojus.;
- *ModelVersion* – tai esybė, kuri yra skirta saugoti modelio versijų informaciją.



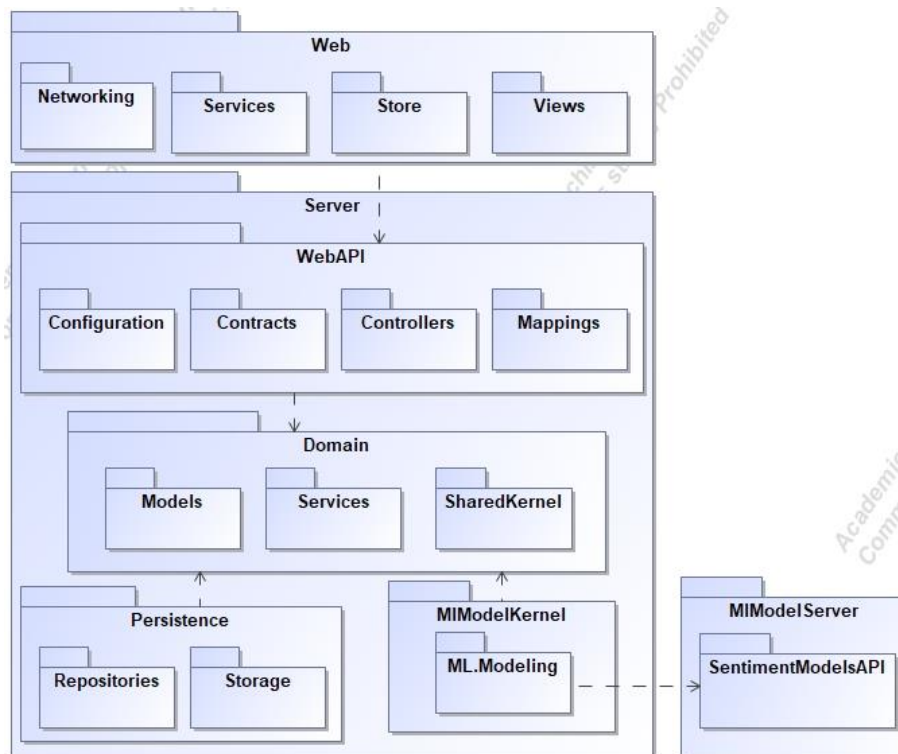
2.4.1 pav. Sentimentų analizės sistemos esybių ryšių diagrama

2.4.2. Statinis sistemos vaizdas

Šiame skyriuje trumpai pateikiamas apibendrintas sistemos statinis paketų vaizdas bei trumpai detaliau aprašomi tik du pagrindiniai sistemos moduliai. Daugiau detalizuojami tik vartotojo sąsajos bei sistemos serverio dalies pagrindiniai komponentai.

2.4.2.1. Bendra sistemos apžvalga

Bendra sistemos architektūra yra sudaryti iš trijų pagrindinių paketų posistemų – vartotojo sąsajos, serverinės sistemos dalies bei modelių valdymo posistemės. Apibendrinta šių posistemų paketų diagrama yra pateikiama 2.4.2 pav.:



2.4.2 pav. Bendra sistemos paketų diagrama

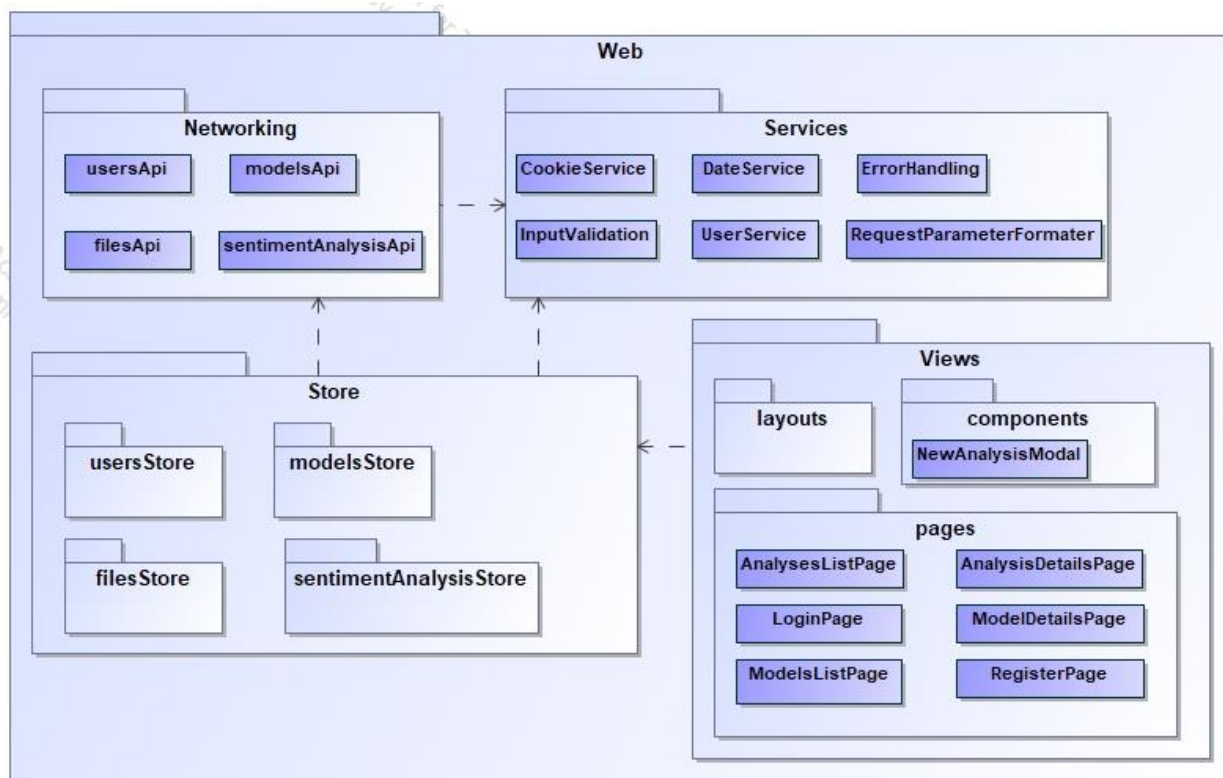
Projektuojamos sistemos programinis kodas bus taip pat atskirtas į tris pagrindines dalis: vartotojo sąsajos, serverinės dalies bei modelių valdymo posistemės ir atskirą programinį. Programinio kodo realizacijai pasirinkta:

- vartotojo sąsajos daliai – *React* biblioteka bei *TypeScript* programavimo kalba;
- sistemos serverinei daliai – *ASP.NET 6* karkasas bei *C#* programavimo kalba;
- natūralios kalbos giliuoju mokymu paremtas modelis – *Flask API* karkasas bei *Python* programavimo kalba.

2.4.2.2. Vartotojo sąsajos paketų detalizavimas

Kliento dalies posistemės detalizuota paketų diagrama yra pateikiama 2.4.3 pav. Diagramoje nėra pateiktos visos posistemėje aprašytos sistemos klasės. *Web* paketas sistemoje yra atsakingas už vartotojo sąsajos programinio kodo išdėstymą. Ši posistemė sukurta naudojantis *React* karkasu, todėl posistemės diagramų paketuose matomas *React Redux* bibliotekų panaudojimo išdėstymas. Toliau pateikiamas atvaizduotų paketų poreikis ir struktūra:

- *Networking* – paketas atsakingas už komunikaciją bei duomenų užklausų suformavimą ir vykdymą tarp vartotojo sąsajos bei sistemos serviso posistemų;
- *Services* – yra pagalbinių klasių paketas;
- *Store* – pakete aprašomos svetainės būsenos išlaikymui reikalingos klasės;
- *Views* – vartotojo sąsajos vaizdui sukurti naudojamos klasės apimančios langų, komponentų bei vartotojo elementų išdėstymą, redagavimą bei atvaizdavimą.

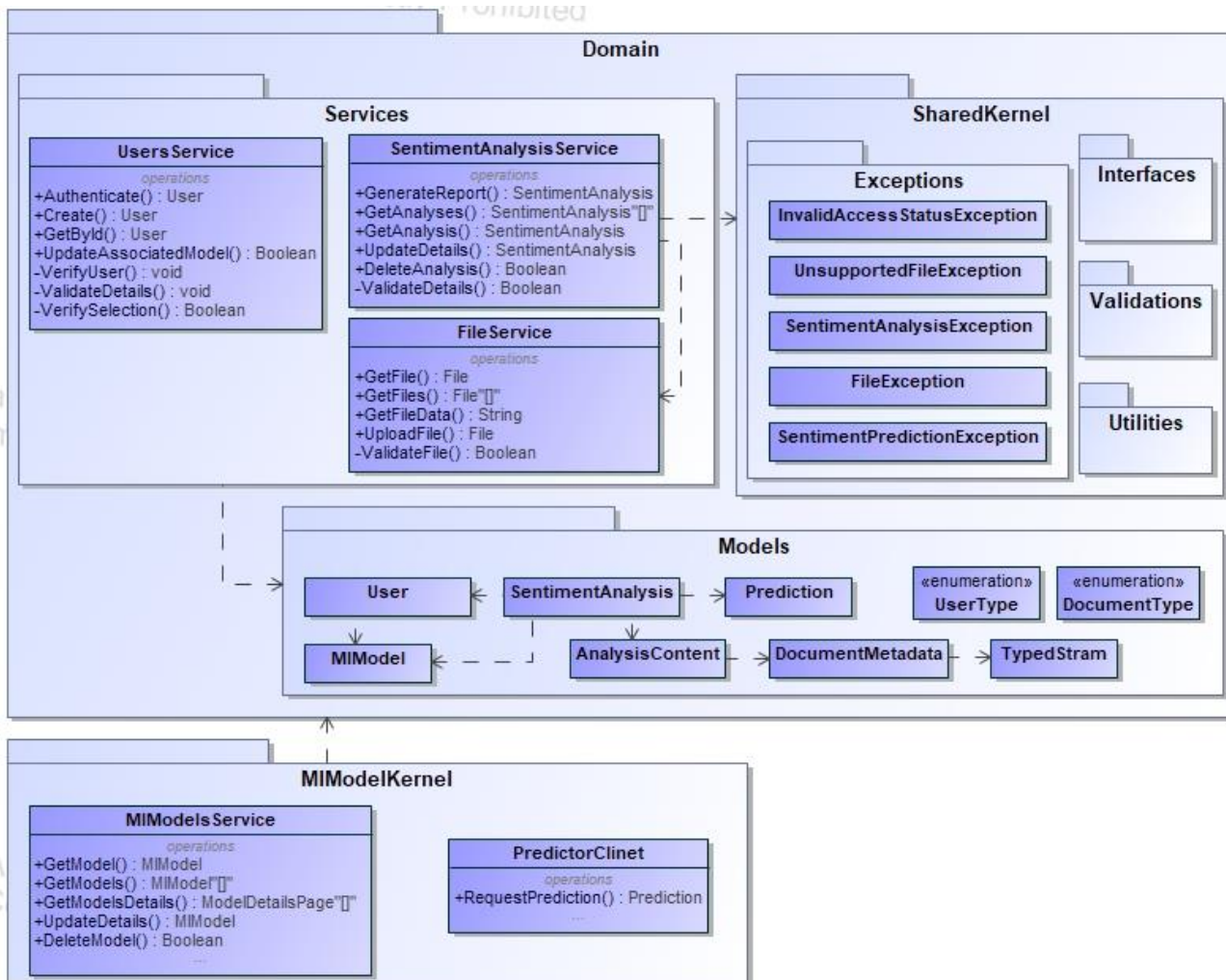


2.4.3 pav. Vartotojo sąsajos posistemės detalizuota paketų diagrama

2.4.2.3. Serverio sąsajos pagrindinės srities paketo detalizavimas

Serverinės dalies *Domain* paketo detalizavimas yra pateikiamas 2.4.4 pav. Šio paketo paskirtis yra apdoroti pagrindinę sistemos užklausų logiką, užtikrinti reikalingas logines bei verslo reikalavimų taisykles, įgyvendinti pagrindinį produkto funkcionalumą. Šis paketas apima pagrindinės sistemos srities taisyklių bei veiksmų užtikrinimą. Vidinių paketų poreikis ir struktūra:

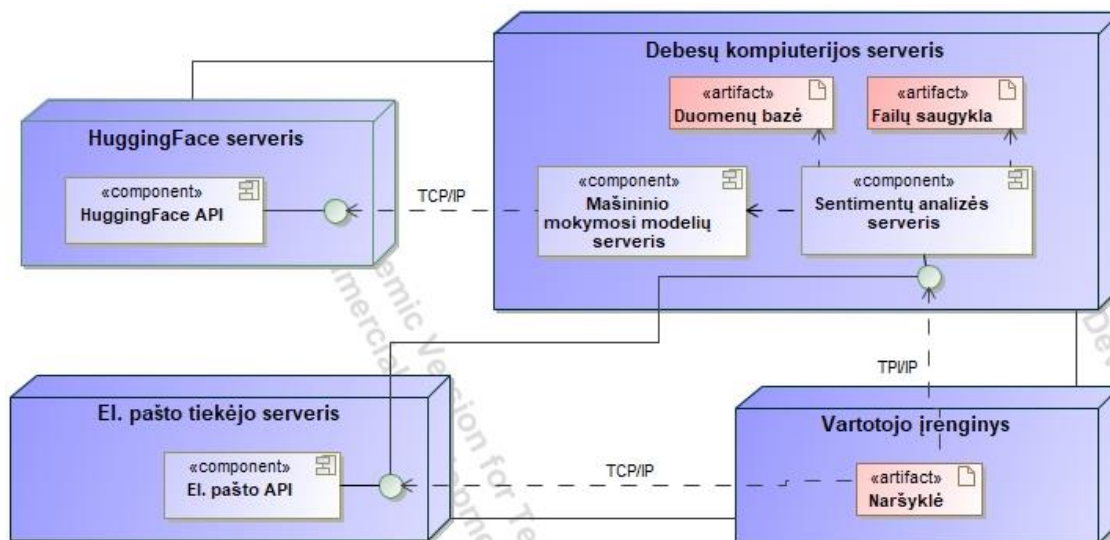
- *Services* – pakete laikomos ir aprašomos operacijos apimančios vartotojų, sentimentų analizės įrašų bei duomenų failų apdorojimą;
- *SharedKernel* – paketas skirtas pagalbinių klasių, metodų plėtinių ir sistemos klaidų objektams aprašyti;
- *Models* – paketas skirtas apibrėžti pagrindinius dalykinės srities modelius;
- *MLModelingKernel* – paketas skirtas mašininio mokymo modelių valdymui, analizių atlikimui bei komunikacijai su sentimentų analizės posisteme.



2.4.4 pav. Serverinės sistemos dalies posistemės detalizuota *Domain* paketo diagrama

2.4.3. Sistemos išdėstymo vaizdas

2.4.5 pav. pavaizduotoje diegimo diagramoje pateikiamas sistemos išdėstymas ir sąryšiai tarp komunikuojančių posistemių ir aplinkinių sistemų.



2.4.5 pav. Sistemos aplinkos diegimo diagrama

Vartotojo įrenginys pasiekia sistemą interneto naršyklėje. Pagal vartotojo poreikius, paskyrą galima susieti su palaikomu el. pašto tiekėjo sistema. Komunikuojančių mazgų atžvilgiu, visi mazgai yra susiję per TCP/IP protokolą. Tai yra standartinis būdas duomenų perdavimui internete. Pateiktoje diegimo diagramoje yra keturi pagrindiniai sistemos komunikacijos mazgai:

- *debesų kompiuterijos serveris* – jame yra patalpinamas pagrindinis sistemos programinis kodas, duomenų bazė ir failų saugykla. Duomenų bazė ir failų saugykla yra fiziniai informacijos saugojimo objektai, o programa yra įdiegiama sistemos dalis;
- *el. pašto tiekėjo serveris* – atstovauja išoriniam el. pašto paslaugų tiekėjui;
- *HuggingFace serveris* – atstovauja išorinei *HuggingFace* dirbtinio intelekto modelių prieigos suteikimo paslaugai. Šie modeliai pasiekiami per programinę sąsają;
- *vartotojo įrenginys* – tai fizinis mazgas, kuriame yra vartotojui pasiekiamą naršyklė.

2.5. Sistemos realizacija

Darbo metu realizuota internetinė svetainė, įgyvendinanti aprašytus panaudojimo atvejus. Sistemos funkcionalumas yra orientuotas į sentimentų analizės modelių išbandymą. Sistemoje vartotojas gali pateikti atsiliepimų įrašus tekstiniu formatu arba įkelti .txt, .csv bei .json tipo dokumentuos. 2.5.1 pav. – 2.5.2 pav. pateikiami dokumentų pavyzdžiai, kurie yra tinkamo formato ir gali būti naudojami tolesniame apdorojime sentimentų analizės sistemoje.

```
{
  "comments": [
    {
      "Comment": "Labai idomus spektaklis"
    },
    {
      "Author": "Anonymous",
      "Comment": "Manau, kad sekantį kartą dar kartą iš čia užsisakinėsiu"
    },
    {
      "CommentName": "Nerekomenduoju!!!",
      "Author": "vartotojas@email.com",
      "Comment": "Baisus aptarnavimas, prasta prekių kokybė, neišmanoma nieko grąžinti!",
      "Date": "2023-09-16"
    }
  ]
}
```

2.5.1 pav. Tinkamai suformatuoto .json dokumento tipo pavyzdys

CommentName	Date	Author	Comment
	2023-09-12		Labai patiko šis restorano meniu bei aptarnavimas
Rekomenduoju suknel	2023-09-11	lneta	Kokybiška medžiaga, labai puikios spalvos.
Puikus restoranas	2024-01-08	Anonymous	Nepatiko skonis, mėsa buvo sausa, garnyro trūko. Padavėjas retai p

2.5.2 pav. Tinkamai suformatuoto .csv dokumento tipo pavyzdys

Sistemoje visi pateikti dokumentai ar tekstiniai duomenys yra toliau konvertuojami į .csv tipo dokumentus. Atliktos analizės rezultatai yra suformuojami ir išsaugomi rezultatų faile. Sistemos vartotojai gali peržiūrėti savo atliktų analizių sąrašą, atitinka atlikti paiešką, filtravimą, rikiavimą pagal pasirinktus kriterijus (žr. 2.5.3 pav.).

Ieškoti pagal pavadinimą arba data

Analizės data ↓	Pavadinimas	Modelis	Statusas	Teigiami komentarai	Failas
2024-04-11	Sentimentų analizė gėlių parduotuvei	Sentimentų analizė - DistilBert lithuanian-sentiment-analyser-001	● Sėkminga	52.60%	Rezultatai_rez
2024-04-11	Kavinės SkanuSkanu atsiliepimai	Sentimentų analizė - T5 t5-senti-decoder	● Laukiama		Nėra rezultatų failo

Įrašų kiekis 10 ▾ 1-2 of 2 < >

2.5.3 pav. Sentimentų analizės sistemos vartotojo analizių sąrašo lango pavyzdys

2.5.4 pav. pavaizduotas atliktos analizės įrašo informacinis langas. Vartotojas gali matyti apibendrintą informaciją, kaip pozityvių ir neigiamų komentarų pasiskirstymas, bendras įrašų kiekis, analizę atlikusio modelio pavadinimas, versija, įvesties bei rezultatų dokumentai. Vartotojui sugeneruojama rezultatų histograma pavaizduota 2.5.5 pav. Sukurtame rezultatų faile pateikiami detalūs prognozavimo rezultatai bei įverčių tikėtino balas.

Home / Mano analizės

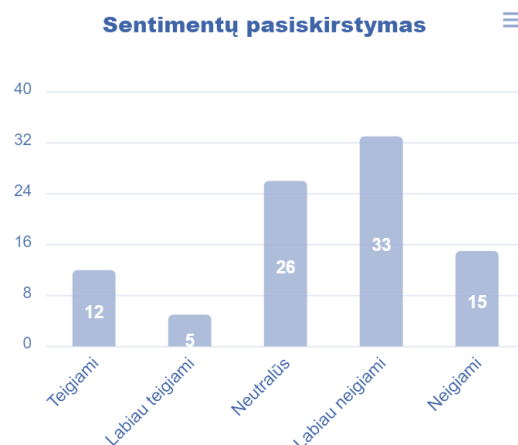
Mano analizės

Sentimentų analizės įrašo informacija

[Turite klausimų?](#) Grįžti į sentimentų analizių sąrašą

Pavadinimas	Sentimentų analizė gėlių parduotuvei		
Sentimentų analizės modelis	Sentimentų analizė - DistilBert lithuanian-sentiment-analyser-001	Populiariausias sentimentas positive	Aanalizės data 2024-04-11
Statusas	● Sėkminga	Iš viso įrašų 91	
Duomenų failas:	Rezultatai.csv	Teigiami bei neigiami komentarai	
Rezultatų failas	Rezultatai_rez.csv	52 %	18 %

2.5.4 pav. Atliktos sentimentų analizės pavyzdys



2.5.5 pav. Atliktos sentimentų analizės prognozių kategorijų pasiskirstymo pavyzdys

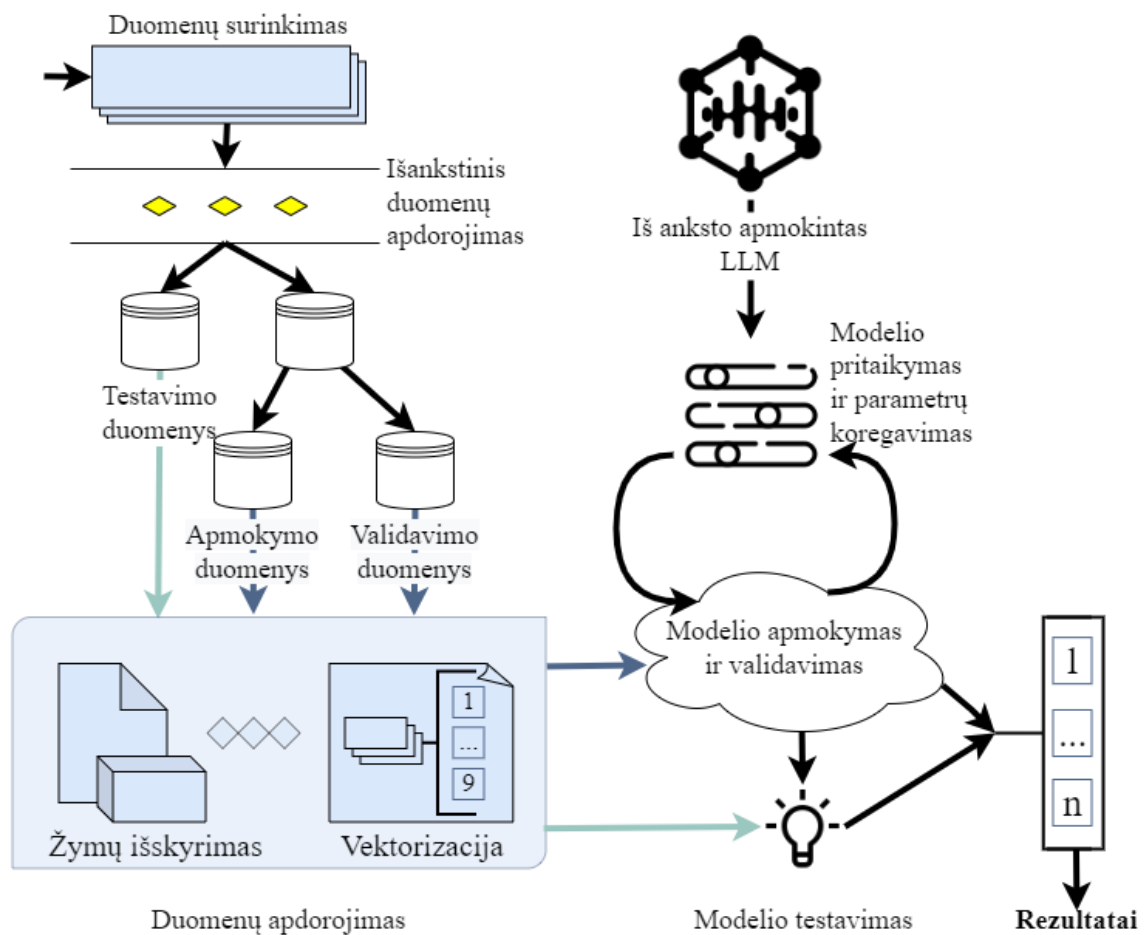
Sistema suteikia vartotojams galimybę išbandyti įvairius sentimentų analizės modelius. Ji apdoroja įvestis įvairiais formatais, tačiau rezultatai visada pateikiami vienodu būdu. Analizės atlikimui nėra apribojimų: naudotojai gali atlikti neribotą skaičių analizių su sistemoje esančiais modeliais. Gautus rezultatus galima peržiūrėti tiesiogiai internetinėje svetainėje arba atsisiųsti juos kaip galutinį failą. Be to, prireikus sistema suteikia galimybę palyginti rezultatus pagalbinais įrankiais.

3. Tyrimo metodai

Šiame skyriuje detalizuojamas sentimentų analizės modelio įgyvendinimas. Aprašomi atliktos veiklos ir detalizuojama apmokyto modelio konfigūracija, įtraukiant duomenų surinkimo, paruošimo procesus, modelio pasirinkimą ir parametrizavimą.

3.1. Modelio kūrimo procesas

Darbo metu siekiama sukurti giliuoju mokymusi paremtą modelį lietuviškų tekstų sentimentų klasifikavimui. Šiuo metu nėra žinoma, kad būtų atlikti LLM lietuvių kalbos sentimentų analizės uždavinio tyrimai, taip pat trūksta išplėtotų lietuvių kalbos sentimentų žodynų. Tyrimo metu bandoma pritaikyti daugiakalbį LLM modelį lietuviškiems tekstams, siekiant pasiekti kuo geresnius tikslumo įverčius. Darbo metu atliekama penkių kategorijų sentimentų analizė. Modeliai mokomi ir testuojami naudojant tyrimo metu surinktą duomenų imtį. Surinkti duomenys yra tik minimaliai apdoroti ir paruošiami LLM mokymui. Šis procesas apriboja galimybę gautus rezultatus tiksliai palyginti su kitų tyrėjų aprašytais tyrimais.



3.1.1 pav. Apibendrinta veiksmų seka sentimentų analizės modelio sukūrimui

Apibendrinta darbo eiga kuriant sentimentų analizės modelį yra pateikiama 3.1.1 pav. Modelio kūrimo etapai:

- duomenų surinkimas – reikalavimų internetinėms svetainėms surinkimas, svetainių paieška, automatizuotų metodų sukūrimas ir pritaikymas atsiliepimų surinkimui;

- duomenų apdorojimas – apima veiksmus, kaip nereikalingų skyrybos ženklų, nereikšmingų jungtukų, simbolių tekste pašalinimas;
- duomenų paskirstymas į testavimo, validavimo ir apmokymo imtis;
- modelio paruošimas – modelio pasirinkimas, pirmųjų parametrų pasirinkimas;
- duomenų paruošimas pasirinktam modeliui – žymų išskyrimas, vektorizacija ir kategorijų paruošimas;
- modelio apmokymas, rezultatų analizė, galimas parametrų koregavimas ir proceso pakartojimas.

Tolesniame apraše išdėstyti detalūs veiksmai, atlikti duomenų rinkimo, paruošimo, modelio parametrų nustatymo ir kitų etapų metu. Abiem analizuojamiems modeliams taikomi vienodi duomenų apdorojimo etapai.

3.2. Duomenų imtis

Darbo metu atliekamas tyrimas skirtas išanalizuoti didelių kalbos modelių panaudojimo galimybes sentimentų analizėje su lietuviškų komentarų tektais. Šiuo metu yra skirtingų būdų, kaip siūloma apmokinti sentimentų analizei modelius, kada duomenų imtys specifinėje kalboje yra nedidelės arba jų nėra išvis. Tačiau dėl darbo temos pasirinkimo, darbo metu nėra naudojama galimybė versti komentarus ir vertinti jų atitikmenis jau apmokytuose kalbos modeliuose. Taip pat, internete nebuvo rasta laisvai prieinamų lietuvių kalbos duomenų bazių pritaikytų sentimentų analizei. Šiame poskyryje pateikiama informacija apie darbo metu surinktas duomenų imtis, jų apdorojimą ir paruošimą apmokymui.

3.2.1. Reikalavimai duomenims

Siekiant išanalizuoti, kaip modelis geba identifikuoti lietuvių kalbos emocijų perteikimo ypatybes, pasirinkta tekstus klasifikuoti į penkias kategorijas. Renkant tinkamus atsiliepimus internete, buvo nustatytas reikalavimas, kad atsiliepimai turėtų penkių balų arba penkių žvaigždučių įvertinimą. Šis duomenų tipas leidžia išvengti papildomo rankinio kategorijų priskyrimo, kadangi etikečių priskyrimas (angl. *labeling*) gali būti subjektyvus. Minėti įvertinimai toliau normalizuojami į vienodo tipo kategorijas pateiktas 3.2.1 lentelėje.

3.2.1 lentelė. Atsiliepimų įvertinimų priskyrimas sentimentų kategorijoms

Įvertinimas arba žvaigždučių kiekis	Sentimentas	Kategorijos reikšmė	
		Skaitinė	Tekstinė
5	emocijomis paremtas teigiamas sentimentas	4	<i>positive</i>
4	racionalumu paremtas teigiamas sentimentas	3	<i>mostly_positive</i>
3	neutralus ar emocijos neperteikiantis tekstas	2	<i>neutral</i>
2	racionalumu paremtas neigiamas sentimentas	1	<i>mostly_negative</i>
1	emocijomis paremtas neigiamas sentimentas	0	<i>negative</i>

3.2.2. Duomenų šaltiniai

Darbo metu duomenys buvo renkami iš viešai prieinamų internetinių svetainių. Duomenų rinkimui internete (angl. *data scraping*) buvo naudojami automatizuoti sprendimo būdai. Priklausomai nuo duomenų šaltinio, rinkimo programinis kodas buvo parašytas šiomis programavimo kalbomis:

- *Python* – scenarijų rašymui naudotas *Selenium* paketas;
- *PHP* – scenarijų rašymui naudotas *PHP-Webdriver* paketas.

Renkantis internetines svetaines atsiliepimų rinkimui, buvo keliami tokie reikalavimai:

- šaltiniai pasiekiami viešai, nereikalaujant registruotos vartotojo paskyros;
- svetainėje palikti atsiliepimai turi būti lietuviški;
- atsiliepimai turi būti vertinami penkiabalėje vertinimo sistemoje (pvz.: *balai, žvaigždutės*).

Remiantis pateiktais reikalavimais bei ieškant pakankamai vartotojų atsiliepimų turinčių svetainių, apmokymo duomenys buvo surinkti iš 3-ijų skirtingų šaltinių. Šaltinių nuorodos bei pagrindiniai paaiškinimai pateikiami 3.2.2 lentelėje.

3.2.2 lentelė. Duomenų rinkinių šaltinių informacija

Svetainės pavadinimas	Svetainės tipas	Surinktų įrašų kiekis	Nuoroda
<i>Pigu.lt</i>	Vienas didžiausių elektroninės prekybos portalų Lietuvoje, siūlantis platų įvairių prekių asortimentą.	5993	https://pigu.lt/lt/
<i>Atsiliepimai.lt</i>	Interneto svetainė, kurioje vartotojai dalinasi savo patirtimi apie įvairias prekes, paslaugas, įmones Lietuvoje.	3212	https://atsiliepimai.lt/
<i>Google Maps</i>	Google sukurtas žemėlapių tinklalapis, kuriame pateikiama informacija apie įvairius objektus. Vartotojai gali vertinti įvairius objektus: parduotuves, restoranus, paslaugas ir kita.	122795	https://www.google.com/maps

Pasirinktų duomenų šaltiniai apima įvairias gyvenimo sritis, įskaitant teikiamų paslaugų kokybės vertinimus, produktų atsiliepimus ir išgyvenimų pasidalinimus. Didžiausia duomenų imtis buvo gauta iš *Google Maps* svetainės. Šie vartotojų atsiliepimai yra surinkti iš įvairių gyvenimo sričių, apimančių, bet neapsiribojant: restoranus, greito maisto užkandines, parduotuves, prekybos centrus, lankytinas vietas ir pramogų centrus.

3.2.3. Duomenų paruošimas ir analizė

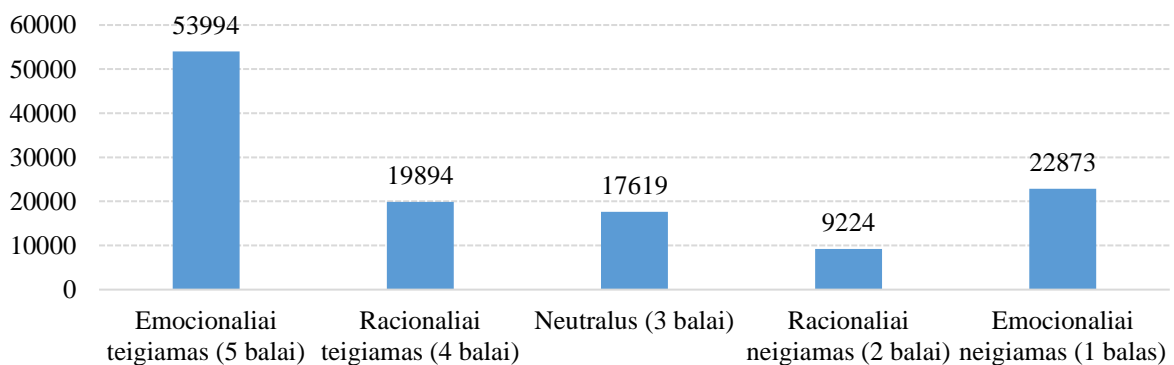
Prieš atliekant duomenų įvertinimą bei sentimentų analizės modelio kūrimą, surinkti duomenys buvo apdoroti, sutvarkyti ir išfiltruoti. Pradinis duomenų apdorojimo procesas vykdomas šia tvarka:

1. *Duomenų anonimizavimas*. Sentimentų modelio apmokymui buvo išsaugomi ir paliekami tik vartotojų atsiliepimai bei jų įverčiai. Visi kiti duomenys apie vartotojus, atsiliepimus, temas buvo pašalinami ir toliau analizės bei mokymosi etapuose nenaudojami. Šis procesas yra svarbus

siekiant naudoti duomenis mokymui ir analizei, užtikrinant asmenų privatumą bei sumažinant modelio šališkumą.

2. *Duomenų sujungimas.* Visi duomenys surinkti iš skirtingų šalininių yra apjungiami ir tolimesnis apdorojimas vyksta visai duomenų imčiai.
3. *Duomenų rinkinio išvalymas.* Pašalinami netikslingi, tušti ar nežodiniai komentarai, pavyzdžiui, komentarai, kurie yra tušti arba sudaryti tik iš jaustukų (angl. *emoji*) simbolių. Taip pat buvo pašalinti įrašai, ilgesni nei 400 simbolių, dėl taikomų modelių konteksto ilgio apribojimų. Bendrai šių reikalavimų neatitinkančių įrašų buvo sąlyginai nedaug, nes ieškant įrašų internete buvo atsižvelgiama į šiuos kriterijus, tačiau buvo pašalinta daugiau nei 700 įrašų.
4. *Ne lietuvių kalbos atsiliepimų išrinkimas.* Visi atsiliepimai buvo papildomai suklasifikuoti pagal kalbą naudojantis Python '*langid*' biblioteka. Atskirti lietuvių kalbos atsiliepimai nuo kitų kalbų tekstų. Buvo suklasifikuota per 50000 ne lietuvių kalbos tekstų. Šie įrašai toliau buvo peržiūrėti automatizuotu būdu, naudojant GPT-4 programinę sąsają, kuri pasižymi geresniu kalbos atpažinimu. Iš '*langid*' bibliotekos suklasifikuotų rezultatų daugiau nei 20000 tekstų buvo priskirti lietuvių kalbai. Likę duomenų rinkiniai buvo peržiūrėti rankiniu būdu, ir neteisingai suklasifikuoti lietuviški tekstai taip pat buvo perkelti į pagrindinį duomenų rinkinį. Trumpi ir aiškūs atsiliepimai, jei buvo galima, buvo išversti ir grąžinti į pagrindinį duomenų rinkinį. Apie 10 000 įrašų buvo ilgesni nei 5 žodžiai, įvairiomis kalbomis. Siekiant nenaudoti automatinių vertėjų ir neprarasti lietuviškos leksikos subtilybių, šie įrašai buvo išmesti iš galutinio duomenų rinkinio.
5. *Duomenų sujungimas ir paruošimas mokymui.* Visi duomenų rinkiniai yra surenkami ir sujungiami į vieną galutinį rinkinį. Pašalinamos nereikalingos, tarpinių skaičiavimų reikšmės, o įvertinimams apibrėžiamos galutinės klasės.

Galutinis duomenų rinkinys sudarytas iš 123604 įrašų. Duomenų įrašų pasiskirstymas pagal sentimentų kategorijas pateikiamas 3.2.1 pav.

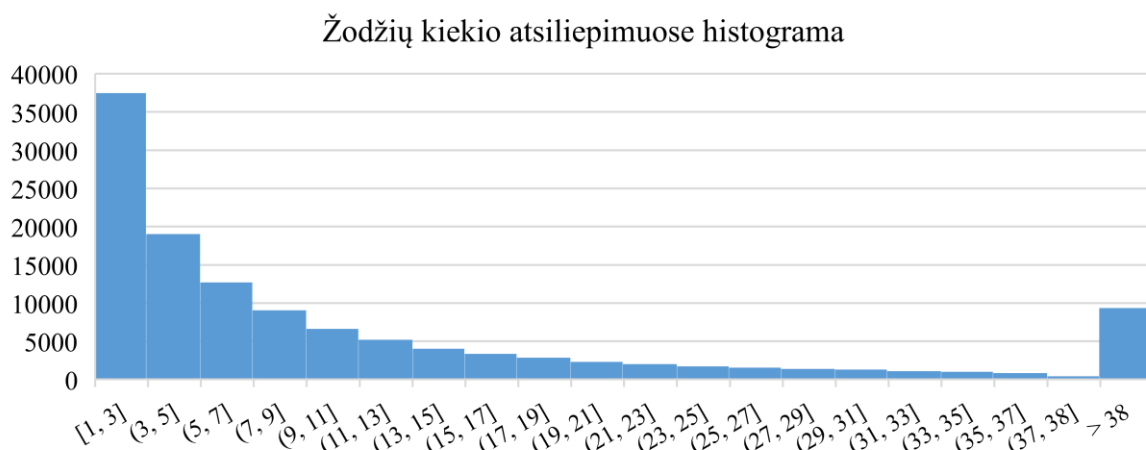


3.2.1 pav. Duomenų įrašų pasiskirstymas pagal 5-ias sentimentų kategorijas

3.2.1 pav. pateikti duomenys yra išsidėstę pagal praktikoje dažnai aptinkamą pasiskirstymą, kai žmonės yra linkę dalintis savo labai puikia arba stipriai prasta patirtimi. Toks požymis atsivaizduoja ir galutinėje duomenų imtyje, kur daugiausia duomenų įrašų turima emocionaliai teigiamoje bei emocionaliai neigiamoje kategorijose. Teigiami sentimentų įrašai sudaro 43% duomenų rinkinio, o neigiami 18.5%.

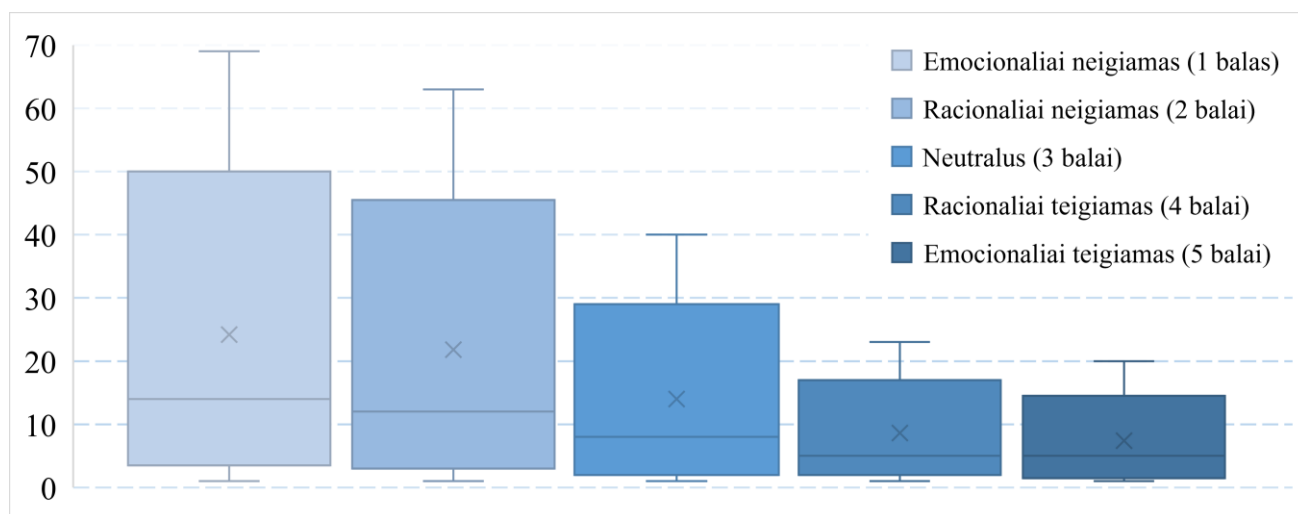
Analizuojant duomenis, matyti, kad didžiausią dalį duomenų rinkinio sudaro trumpi atsiliepimai nuo 1 iki 11 žodžių. Tokios apimties tekstai dažniausiai susideda iš trumpų komentarų, išsireiškimų ir

retai būna ilgesni nei dviejų sakinių. Žodžių kiekio pasiskirstymo histograma duomenų imtyje yra pateikiama 3.2.2 pav.



3.2.2 pav. Žodžių kiekio pasiskirstymo histograma duomenų imtyje

3.2.2 pav. pateiktoje histogramoje atsiliėpimų tekstai, kurie yra ilgesni nei 38 žodžiai sudaro vos 7.5% visos duomenų imties. Remiantis šia histograma, galima teigti, kad didžioji dalis duomenų imties sudaryta iš gana struktūruotų ir trumpų išsireiškimų. Bendrai analizuojant žodžių kiekį atsiliėpimų sąrašė, galima atkreipti dėmesį į skirstinių reikšmes pagal sentimentų kategorijas. Grupių sklaidos pagal turimų žodžių kiekį diagrama pateikiama 3.2.3 pav.



3.2.3 pav. Sentimentų kategorijų tolydinio skirstinio reikšmių diagrama

Diagramoje pateikiamos minimalių ir maksimalių kategorijų žodžių kiekių reikšmės atspindi žemiausius ir aukščiausius duomenų taškus pagrindiniame duomenų diapazone aplink ketvirčius, neįtraukiant išskirtinių ekstremumo atvejų. Duomenų imtyje yra mažiau nei 0.5% atsiliėpimų sudarytų iš daugiau nei 150 žodžių. Taip pat, diagramoje galima matyti, kad didžiausia žodžių kiekio sklaida atsispindi emocionaliai ir racionaliai neigiamų komentarų kategorijose. Bendriausiuose atvejuose, teigiamų komentarų pogrupiuose pasiskirstymas yra tankiausias ir artimiausias normaliajam pasiskirstymui, matoma mažiausia žodžių sklaida. Ilgiausias neigiamas komentaras sudarytas iš 357 žodžių, o ilgiausias teigiamas iš 300 žodžių. O bendras žodžių vidurkis per visą apmokymo duomenų imtį yra 13,2.

Visi šie duomenys toliau buvo suskirstyti į tris duomenų rinkinius: apmokymo, validacijos bei testavimo. Kadangi duomenų kategorijų pasiskirstymas yra didelis, modelio kūrimo metu buvo analizuojamas ir bandomas dviejų skirtingų pasiskirstymų duomenų rinkinių efektyvumas.

3.2.3 lentelė. Duomenų pasiskirstymas apmokymo, validacijos bei testavimo duomenų imtyse

Parametrai Duomenų grupė	Duomenų kiekis			Kategorijų pasiskirstymas, %				
	Apmokymas	Validacija	Testavimas	5	2	3	2	1
Pirma	84050	24721	14833	43,6	16,1	14,2	7,4	18,5
Antra	55489	18497	18497	24,7	21,5	19,1	9,9	24,7

Lentelėje pateikiamų testavimo duomenų imties pasiskirstymas pasirinktas, siekiant:

- pirmuoju atveju, apmokyti modelį ant realaus pasiskirstymo duomenų rinkinio, sutinkamo praktikoje;
- antruoju atveju, apmokyti modelį su šiek tiek standartizuotu duomenų rinkiniu, siekiant gauti geresnius klasifikavimo rezultatus.

Antroje duomenų imtyje, bendras duomenų kiekis yra sumažinamas iki 92483 įrašų. Taip pat, siekiant išlaikyti teigiamų įrašų pasiskirstymą ir įvairiapusiškumą, atliekant duomenų atrinkimą, buvo atsižvelgta ir siekta pašalinti pasikartojančius ir persidengiančius teigiamus įrašus. Kitas rekomenduojamas būdas, kuris yra sutinkamas praktikoje – tai mažiau įrašų turinčių kategorijų papildymas modifikuojant jau turimus įrašus. Šis metodas nebuvo pasirinktas dėl laiko stokos ir laisvai prieinamų lietuviškos kalbos sinonimų žodynų trūkumų.

3.3. distilBERT lietuviškų atsiliepimų sentimentų analizėje

Šiame skyriuje sentimentų analizės uždaviniui spręsti pasitelkiant LLM modelius analizuojamas DistilBERT tipo modelis. DistilBERT modeliai yra 40% mažesni, 60% greitesni nei BERT, tačiau sugeba išsaugoti 97% BERT tikslumo [56]. Šie modeliai dažnai pasirenkami vietoje pagrindinių BERT tipo modelių, kaip pigesni, greičiau apmokomi bei mažiau resursų pareikalaujantys modeliai. Be to, šių modelių kategorijos yra geriau pritaikytos užduotims, kurių apmokymui yra naudojama riboto dydžio duomenų imtis. Toliau analizuojamas bei koreguojamas *'distilbert-base-multilingual-cased'* modelis.

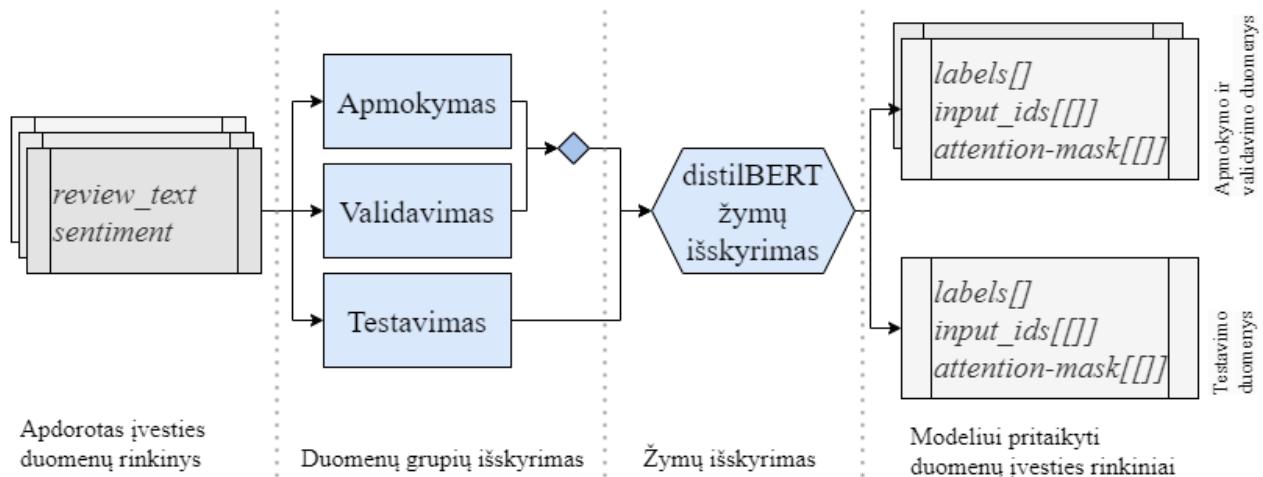
'distilbert/distilbert-base-multilingual-cased' modelis yra supaprastinta BERT modelio versija. Distiliuota modelio versija yra greičiau veikianti bei mažesnės apimties. Modelis išankstinio apmokymo metu buvo apmokytas *Wikipedia* svetainės duomenimis 104 kalbomis. Viena iš apmokymo kalbų buvo lietuvių kalba [39]. Ši modelio versija taip pat pasižymi didžiųjų ir mažųjų raidžių jautrumu. Tai reiškia, kad modelis geba atskirti teksto skirtumus tarp *'labai nepatiko'* ir *'LABAI nepatiko'* įvesčių. Modelis sudarytas iš 6 sluoksnių, 768 dimensijų bei 12-os skirtingų dėmesio sutelkimo mechanizmų [56]. Modelis iš viso yra sudarytas iš 134 milijonų parametru.

Modelio pritaikymo lietuvių kalbos sentimentų analizei metu buvo analizuoti įvairūs koregavimo ir optimizavimo metodai. Toliau detaliau aprašomas modelis, parodęs geriausius rezultatus, jo koregavimo metu pasirinkti hiperparametrai, optimizavimo algoritmas ir rezultatų vertinimas.

3.3.1. Duomenų žetonizacija

Prieš pateikiant duomenis modeliui buvo atlikta pirminė tekstinių įvesties duomenų žetonizacija. Naudojamas *WordPiece* žetonizacijos algoritmu paremtas modelis, kuris yra pritaikytas įvairiakalbio distilBERT modelio apmokymui. Išskiriamos žymos, kurios toliau pateikiamos modeliui apmokymo bei rezultatų įvertinimo skaičiavimo metu.

Modelis yra pritaikytas apdoroti įvestis sudarytas iš ne daugiau kaip 512 žymų. Darbo metu palaikomas žymų konteksto kiekis sumažintas iki 400, dėl įvesties duomenų pasiskirstymo pagal žodžių kiekį bei ribotų skaičiavimo resursų.



3.3.1 pav. Apibendrintas įvesties duomenų konvertavimui į modelio formatą procesas

3.3.1 pav. diagramoje pavaizduotas supaprastintas duomenų imties paruošimo procesas. Apdorotas duomenų sąrašas yra transformuojamas iš įvesties teksto bei sentimentų kategorijų į du modeliui priimtino duomenų formato rinkinius. Pirmajame rinkinyje saugomi apmokymo bei validavimo duomenys, o antrajame testavimo duomenys. Galutinis formatas sudarytas iš:

- *labels[]* – tai sentimentų kategorijų žymų sąrašas. Šis sąrašas yra skirtas identifikuoti kiekvieną tekstinės duomenų įvesties klasę apmokant modelį atlikti klasifikavimo užduotis;
- *input_ids[][]* – tai žetonizuotų įvesties duomenų sąrašų rinkinys;
- *attention_mask[][]* – tai dėmesio įverčių sąrašų rinkinys, kuris skirtas nurodyti modeliui į kuriuos įvesties žetonus reikia sutelkti pradinį dėmesį. Šio sąrašas leidžia modeliui atskirti prasmingas *input_ids* sąrašo žymas, kadangi paruošimo metu yra suvienodinami įvesties duomenų ilgiai, ir trumpesni tekstai yra užpildomi iki maksimalaus palaikomo konteksto ilgio.

Galiausiai, siekiant gauti objektyvius modelio prognozavimo rezultatus, testavimo duomenys yra saugomi atskirame duomenų rinkinyje.

3.3.2. Architektūros sluoksniai ir svarbiausi komponentai

Darbo metu '*distilbert/distilbert-base-multilingual-cased*' modelis yra pritaikomas kategorizavimo uždaviniui. Modelio karkasas iš savęs nėra paruoštas kategorizuoti įvesties duomenis. Todėl atliekant tolimesnį modelio koregavimą yra pridedami nauji papildomi komponentai leidžiantis atlikti teksto klasifikavimą. Aukščiausio lygmens modelio blokai:

- ‘*distilBERT*’, sudarytas iš:
 - žymų įterpimo sluoksnio – skirtas įvesties žymų masyvus atvaizduoti iš anksto žinomais modelio žodyno vektoriais;
 - transformerio bloko – esminė architektūros dalis sudaryta iš 6 transformerio blokų ir skirta sudaryti sudėtingų priklausomybių tarp įvairių įvesties žetonų atvaizdavimą.
- ‘*pre_classifier*’ – skirtas atlikti naudingas duomenų transformacijų savybes leidžiančias lengviau identifikuoti įvesties duomenų išskirtines savybes;
- ‘*classifier*’ – skirtas atlikti klasifikavimo sprendimą pritaikant *softmax* funkciją. Šis sluoksnis yra sukuriamas su neinicializuotais parametru svoriais, kurie mokymosi metu yra keičiami pagal atitinkamus rezultatus ir nustatytus hiperparamterus;
- ‘*dropout*’ – skirtas padėti reguliuoti modelio persimokinimą prie duomenų, ypač naudojant nedideles duomenų imtis.

Modelio koregavimo metu pasirinkta koreguoti visų modelio sluoksnių parametrus. Tai suteikia galimybę modeliui geriau prisitaikyti prie specifinės kalbos ypatumų bei palieka daugiau lankstumo parametrų pokyčiams. Ši metodika dažnai pasižymi geresniais rezultatais, ypač koreguojant daugiakalbės kalbos modelius specifinės kalbos uždaviniui, kaip lietuvių kalbos sentimentų analizė.

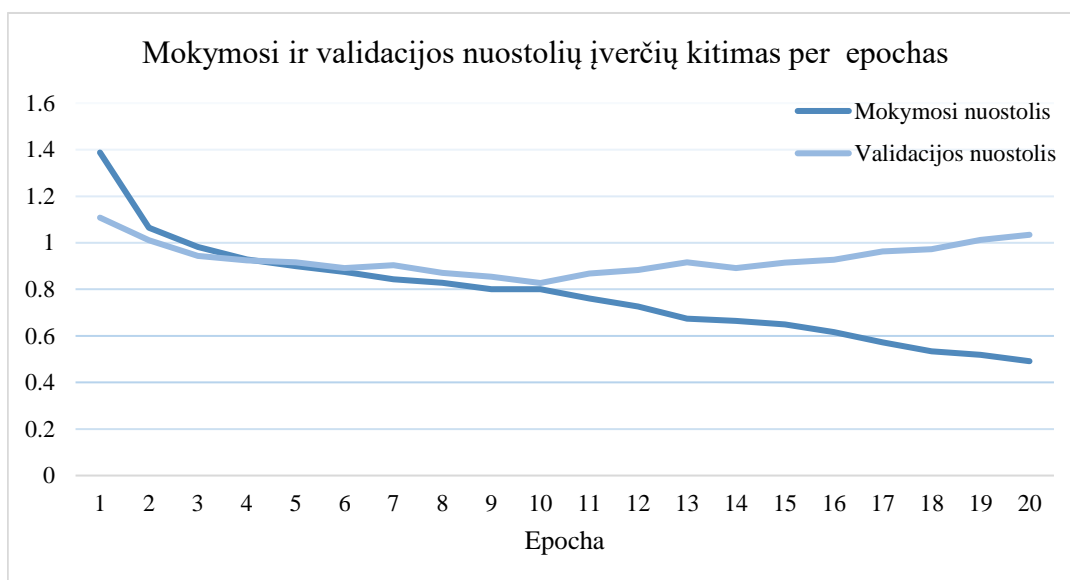
Pagrindiniai modelio koregavimo metu naudojami komponentai:

- *AdamW* optimizavimo algoritmas – tai žinomo Adam optimizavimo algoritmo variantas, kuris pasižymi geresniais apibendrinimo rezultatais. Ši algoritmo variacija atskiria svorių nykimo reguliavimą nuo kinančių vidurkių ir jų kvadratų [57]. Šis pasikeitimas leidžia modeliui efektyviau valdyti svorių atnaujinimą bei geriau apibendrinti duomenis;
- *Softmax* klasifikavimo funkcija – naudojama išvesties sluoksnyje norint atlikti daugiaklasį sentimentų klasifikavimą.

Apibendrinant, pasirinkus papildomus architektūros sluoksnius bei esminius apmokymui reikalingus komponentus modelis yra paruošiamas modelio koregavimo etapui. Architektūros papildymas klasifikavimo bloku suteikia galimybę tolimesnį modelio mokymosi tikslą orientuoti į klasifikavimą.

3.3.3. Hiperparametrai ir rezultatai

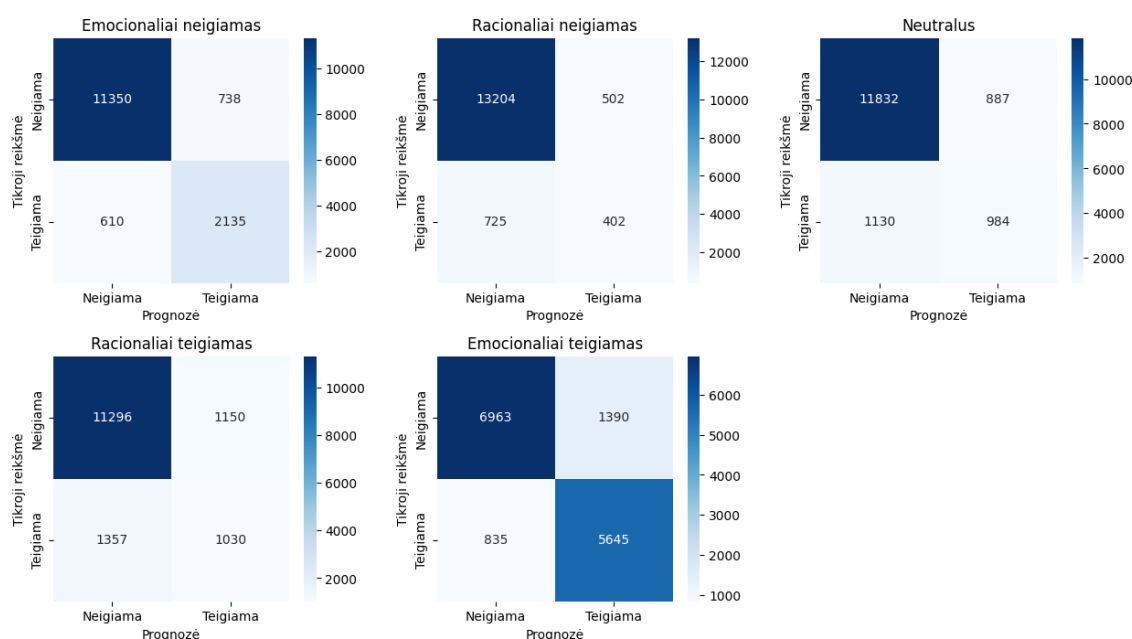
Modelio apmokymo ciklą kokybės vertinimui yra naudojama kryžminė entropija. *distilBERT* modelis buvo apmokytas 25 epochas. Modelio mokymosi ir validacijų nuostolių įverčių kitimas epochose yra pateikiamas 3.3.2 pav.



3.3.2 pav. Pirmojo duomenų rinkinio distilBERT modelio mokymosi nuostolių įverčių kitimo diagrama

Pateiktoje diagramoje galima išvelti greitą mokymosi bei validavimo nuostolių kitimo mažėjimą iki 10-tos epochos. Tačiau tolimesnių epochų nuostolių kreivės pradeda išsiskirti. Galima matyti stabilų mokymosi įverčių mažėjimą, tačiau validavimo etape šis įvertis pradeda įgauti vis didesnes reikšmes nuo 15-tos epochos. Modelis greitai pradeda prisitaikyti prie jam pateikiamų apmokymo imties duomenų.

3.3.3 pav. pateikiamos kiekvienos sentimentų kategorijos maišaties matricos. Diagramose matomi modelio spėjimų ir tikrųjų įverčių klasifikavimo rezultatai. Vertinant šias diagramas, svarbu atsižvelgti į kategorijų disbalansą, kuris gali klaidinti ir rodyti pernelyg gerus neigiamų kategorijų priskyrimo rezultatus.



3.3.3 pav. distilBERT modelio, apmokyto su pirmuoju duomenų rinkiniu, individualių sentimentų kategorijų maišaties matricos

Modelis pasižymi gebėjimu gana gerai atskirti emocionalumu paremtas sentimentų kategorijas. Tačiau pastebimas didesnis klaidų pasiskirstymas tarp racionaliai neigiamų ir teigiamų prognozių. Galima išvelgti modelio negebėjimą atskirti racionalių komentarų. Tai taip pat susiję su duomenų imties pasiskirstymu, kadangi šių kategorijų įrašų duomenų rinkinyje yra mažiausia dalis, modelis gali pradėti atpažinti ir kategorizuoti labiau tikėtinas kategorijas.

3.3.1 lentelėje yra pateikiama dešimtis atsitiktinai atrinktų testavimo duomenų imties įrašų iš visų penkių prognozuojamų sentimentų kategorijų.

3.3.1 lentelė. Atsitiktinai parinktų testavimo įrašų distilBERT modelio prognozių pavyzdžiai

Indeksas	Įvesties duomenys	Tikra kategorija	Modelio prognozė	Teisingas
15618	Tragedija, niekada daugiau negrįšiu.	0	0	Taip
18147	Maistas neblogas tačiau margaritos nebe tokios kaip anksčiau. Tačiau didžiausia problema - aptarnavimas. Mus apšaukė todėl kad užsirezervavę stalą atėjom išgerti tik kokteilių ir užkandžių bet nesisakėm antrų patiekalų . Linkime atsakingiau rinktis personalą kadangi po tokių patirčių grįžti nebesinori.	0	1	Ne
12209	Viskas labai brangu. Įkišo apipuvusius pomidorus už didelę kainą ir braškės indelyje viršuje buvo raudonos apačioje - visiškai neišnokę baltos. Vaikščiodama apžiurejau viduje esančių kavinių kainas: sriuba - 3 eur bulviniai blynai - 6 EUR plovo 600 g porcija - 8 EUR dienos pietūs - 7 EUR. Mėsos žuvies taip pat neapsimoka pirkti nes prekybos centruose tą pati produkcija kainuoja mažiau. Pastaruoju metu buvau kelis kartus ir turgus visada buvo apytustis.	1	1	Taip
17139	Lėtas personalas maistas atėjo prieš alų. Sveiki????! Alus lėčiau nei šiltas maistas.	1	0	Ne
8765	Per didelis...	2	2	Taip
10470	Viena iš nemėgstamiausių Maximos parduotuvių. Visada būna tvankupardavėjos piktos ir nemandagios. Žmonių daug.	2	3	Ne
3874	greitas prekių pristatymas	3	3	Taip
6260	Smagiai ką norėjau tą gavau.Labai malonus aptarnavimas	3	4	Ne
5219	Nuostabi krautuvė yra beveik visko ko skrandis ir širdelė geidžia;)	4	4	Taip
1798	Greitai aptarnavo nereikėjo laukt ilgai nors ir daug žmonių buvo.	4	3	Ne

Pateiktuose pavyzdžiuose galima išvelgti duomenų kokybės išskirtinumus, perteikiamas emocijas, gramatikos, sintaksės klaidas. Visi duomenys yra originalaus formato, kadangi nebuvo taikyti jokie teksto apdorojimo, sintaksės ar lemantizacijos metodai. Tai leidžia modeliui geriau atrasti papildomų ryšių tarp įvairių leksikos, sintaksės bei gramatikos ypatybių. Vertinant šių duomenų kokybę, nesunku pastebėti, kad tam tikri atsiliepimai galėtų būti neteisingai suklasifikuoti ir žmogiškųjų resursų, dėl duomenų dviprasmiškumo, riboto konteksto ir kitų kokybės problemų. Be to, iš pateiktų pavyzdžių bei maišaties rezultatų lentelės, galima matyti, kad net ir neteisingų klasių kategorijų priskyrimas dažnai sumaišomas su gretimomis klasėmis.

4. Tyrimo rezultatai

Magistrinio darbo metu buvo analizuojami bei išbandomi naujausi NLP klasifikavimo uždaviniuose taikomi metodai. Tyrimo metu buvo koreguojami du skirtingų architektūrų transformerio tipų modeliai:

- distilBERT modelis – paremtas koduotuvo architektūra;
- ByT5 modelis – paremtas koduotuvo–iškoduotuvo architektūra.

Taip pat, tyrimo metu buvo analizuojamas GPT-4 modelio gebėjimas kategorizuoti testavimo rinkinio duomenis pasitelkiant užklausų (angl. *prompt*) inžineriją.

distilBERT ir ByT5 modeliai išankstinio apmokymo metu buvo apmokyti su daugiakalbiu duomenų rinkiniu. O šio tyrimo metu, abu modeliai pritaikyti lietuviškų sentimentų analizės klasifikavimo uždaviniui, su 3-čiame skyriuje aprašytais duomenų rinkiniais.

Modelių koregavimas, sentimentų analizės uždavinio apmokymas bei įvertinimo metrikų surinkimas buvo vykdomas *Paperspace*⁹ darbo aplinkoje. Darbo metu naudojama mokamoji A6000 virtuali mašina, turinti 48 GB grafinio procesoriaus atminties ir 45 GB operatyviosios atminties.

4.1. distilBERT

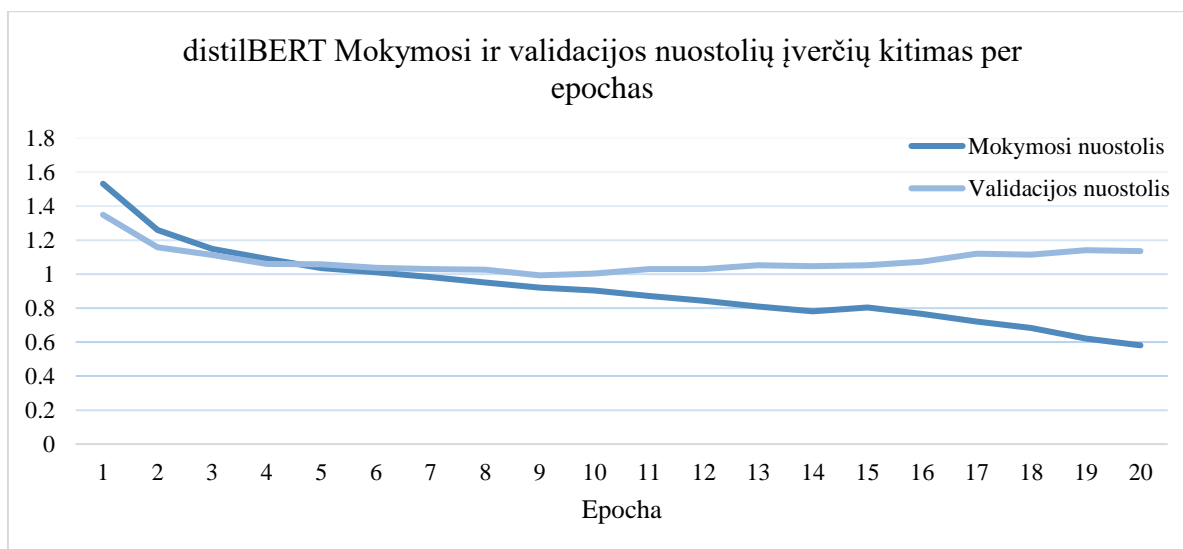
Tyrimo metu buvo sukurti du distilBERT tipo modeliai. Šiame skyriuje detalizuojami du geriausius tikslumo ir F1 įverčio rezultatus parodę distilBERT modeliai.

Bandymų metu buvo pastebėta, kad distilBERT modelis su turimais klasifikavimo duomenimis, greitai prisitaiko prie duomenų imties ir pradeda grąžinti vis prastesnius epochų validavimo rezultatus. Siekiant padidinti modelio gebėjimą apibendrinti duomenų požymius buvo taikomos mokymosi reguliarizacijos technikos, kaip:

- *drop_out* parametrai, kurie nurodė praleisti 30% mazgų bei 35% savipastabos sluoksnių modelio struktūroje, mokymosi metu;
- mažas ir kintantis mokymosi greitis;

3.3.2 pav. ir 4.1.1 pav. pateikiamos diagramos atvaizduojančios distilBERT modelio mokymosi ir validavimo nuostolių kitimą epochose su skirtingais duomenų rinkiniais. Tačiau tiek su didesnės, tiek su mažesnės apimties duomenimis, modelis greitai įsimena mokymosi duomenis. Tačiau su didesniu, pirmuoju duomenų rinkiniu šis procesas ne tik ne toks staigus, bet ir modelis sugeba pasiekti geresnius validavimo rezultatus. Toliau analizuojami modeliai yra parenkami ne tik pagal validavimo nuostolio įvertį, bet ir pagal validavimo duomenų rinkinio tikslumo bei F1 įverčius, todėl taikomas ankstyvo sustabdymo principas ir modeliai atkuriami iš geriausių rezultatų parodžiusių patikros punktų (angl. *checkpoint*).

⁹ Paperspace – debesų kompiuterijos platforma (nuoroda: <https://www.paperspace.com/>)



4.1.1 pav. Antrojo duomenų rinkinio distilBERT modelio mokymosi nuostolių įverčių kitimo diagrama

Testavimo duomenų imties prognozės rezultatai pateikiami maišaties matrica. 4.1.1 lentelėje atvaizduojama maišaties matrica visoms penkioms sentimentų kategorijoms. Lentelė atvaizduoja klasifikavimo rezultatus vertinant modelio spėjimus su tikromis įverčių reikšmėmis. Eilutėse pateikiamos tikrosios sentimentų reikšmės iš duomenų imties, o stulpeliuose – modelio prognozės. Pagrindinėje įstrižainėje tikroji kategorija sutampa su prognozės kategorija. Ši įstrižainė leidžia pamatyti kiekvienos kategorijos teisingai klasifikuotų atvejų skaičių. Ryškiausi ir tikslingiausias kategorijų prognozės matomos klasifikuojant emociniai neigiamus bei emociniai teigiamus sentimentus. Taip pat, lentelėje matoma, kad modelio neteisingai prognozuojamos klasės yra bendrai stipriau koreliuojančios su matricos lentelės įstrižaine. Neutrali bei racionaliai neigiama kategorijos yra pasižymintios didžiausių klaidingai atliktų spėjimų sklaida.

4.1.1 lentelė. distilBERT modelio prognozių maišaties matricos lentelė su pirmuoju duomenų rinkiniu

Prognozės kategorija \ Tikra kategorija	Emocionaliai neigiamas	Racionaliai neigiamas	Neutralus	Racionaliai teigiamas	Emocionaliai teigiamas
Emocionaliai neigiamas	2135 (80.74%)	248 (9.38%)	197 (7.45%)	82 (3.10%)	83 (3.14%)
Racionaliai neigiamas	362 (26.32%)	402 (29.20%)	232 (16.85%)	71 (5.15%)	40 (2.91%)
Neutralus	237 (12.76%)	217 (11.69%)	984 (53.00%)	396 (21.31%)	280 (15.08%)
Racionaliai teigiamas	48 (2.63%)	32 (1.75%)	299 (16.41%)	1030 (56.51%)	978 (53.60%)
Emocionaliai teigiamas	71 (1.14%)	25 (0.40%)	149 (2.37%)	590 (9.39%)	5645 (89.61%)

4.1.2 lentelėje pateikiama modelio sentimentų kategorijų prognozių maišaties matrica, apmokyto su antruoju duomenų rinkiniu. Šioje maišaties lentelėje matomas ryškesnis kategorijų pasiskirstymas įstrižainėje, kadangi šis duomenų rinkinys turi šiek tiek tolygesnį kategorijų pasiskirstymą. Modelis, geriausiai pasižymi emociniai teigiamų bei neigiamų komentarų atpažinime. Bei matomas didesnė spėjimų maišatis tarp neutralių bei racionaliai teigiamų komentarų, dažnai šių atsiliepiamų kategorijos yra supainiojamos tarpusavyje.

4.1.2 lentelė. distilBERT modelio prognozių maišaties matricos lentelė su antruoju duomenų rinkiniu

Prognozės kategorija Tikra kategorija	Emocionaliai neigiamas	Racionaliai neigiamas	Neutralus	Racionaliai teigiamas	Emocionaliai teigiamas
Emocionaliai neigiamas	3474 (80.01%)	489 (11.26%)	375 (8.63%)	138 (3.18%)	99 (2.28%)
Racionaliai neigiamas	624 (33.61%)	616 (33.20%)	429 (23.09%)	126 (6.79%)	50 (2.69%)
Neutralus	411 (13.61%)	322 (10.67%)	1745 (57.81%)	804 (26.64%)	242 (8.02%)
Racionaliai teigiamas	121 (3.58%)	83 (2.46%)	505 (14.94%)	2431 (71.96%)	839 (24.81%)
Emocionaliai teigiamas	122 (2.61%)	43 (0.92%)	220 (4.71%)	928 (19.86%)	3261 (69.90%)

Šios maišaties matricos lentelės padeda vizualiai geriau įvertinti modelio gebėjimą prognozuoti kategorijas. Tačiau pateiktų lentelių rezultatuose gali būti sudėtingiau pastebėti tikrųjų kategorijų pasiskirstymą. Svarbu pastebėti, kad abu duomenų rinkiniai yra sudaryti iš skirtingo kategorijų kiekio pasiskirstymo ir net idealiu sentimentų prognozių atveju, maišaties matricų lentelės įstrižainėje įrašų pasiskirstymas nebūtų tolygus.

4.1.3 lentelėje pateikiami distilBERT, apmokyto su pirmuoju duomenų rinkiniu, prognozių rezultatai trimis sentimentų kategorijoms. Lentelė sudaryta remiantis penkių sentimentų kategorijų klasifikavimo rezultatais. Lentelėje matoma, kad modelis gana gerai prognozuoja neigiamus komentarus (3147 teisingi atvejai) ir puikiai atpažįsta teigiamus komentarus (8243 teisingi atvejai). Tačiau pastebimas duomenų disbalansas – teigiamų komentarų yra žymiai daugiau nei kitų kategorijų, o tai gali įtakoti modelio rezultatus ir netikslumus neutralioje kategorijoje.

4.1.3 lentelė. distilBERT geriausio modelio 3 sentimentų kategorijų prognozių maišaties matricos lentelė

Prognozės kategorija Tikra kategorija	Neigiamas	Neutralus	Teigiamas
Neigiamas	3147 (75.79%)	429 (10.34%)	276 (6.65%)
Neutralus	454 (14.90%)	984 (32.18%)	676 (22.09%)
Teigiamas	217 (2.98%)	445 (6.11%)	8243 (91.01%)

4.1.4 lentelėje pateikiama geriausius rezultatus parodžiusių modelių tikslumo bei našumo apžvalgos. Modelis, kuris buvo apmokintas su didesniu duomenų kiekiu, parodė geresnius rezultatus net ir esant didesniai duomenų disbalansui tarp kategorijų. Taip pat, didesnis duomenų rinkinys leido modeliui apsimokinti didesnę epochų kiekį. Abiejų modelių tikslumo įverčiai tarp validacijos ir testavimo duomenų rodo, kad modelis geba gana gerai apibendrinti turimus duomenų požymius.

4.1.4 lentelė. distilBERT modelių našumo ir tikslumo įverčių apžvalga

Modelis	Duomenų rinkinys	Epochų skaičius	Tikslumas, %		Mokymosi trukmė
			Validacijos duomenų rinkinys	Testavimo duomenų rinkinys	
distilBERT _{BASE}	1	14	68,43	68,41	~ 6 h
distilBERT _{BASE}	2	12	62,01	62,31	~ 4 h

4.1.5 lentelėje pateikiami sentimentų analizės prognozių rezultatai su testavimo duomenis skirtingiems duomenų rinkiniams. Kaip minėta anksčiau, su pirmuoju duomenų rinkiniu gauti geresnio tikslumo rezultatai.

4.1.5 lentelė. distilBERT modelių sentimentų analizės rezultatai

Modelis	Duomenų rinkinys	Atkūrimo įvertis	F1 įvertis					
			Bendras F1 įvertis	Sentimento kategorijos įvertis				
				1	2	3	4	5
distilBERT _{BASE}	1	0,6845	0,6751	0,7601	0,3556	0,4938	0,4513	0,8354
distilBERT _{BASE}	2	0,6201	0,6201	0,7449	0,3626	0,5134	0,5739	0,7194

Abu modeliai prasčiausius rezultatus pateikia kategorizuodami racionaliai neigiamus komentarus. Modeliai gerai geba atpažinti emocionaliai neigiamus ir teigiamus komentarus, ypač aukštas F1 įvertis yra matomas su pirmuoju duomenų rinkiniu. Šiame duomenų rinkinyje didžiausią dalį komentarų sudaro penktoji kategorija. Vis dėlto subalansuotas bendras F1 įvertis, kuris įvertina komentarų pasiskirstymą, pasiekė 0,6751, kas gali būti vertinama kaip ganėtinai aukštas rezultatas.

Galiausiai, pasinaudojus modelio, parodžiusio geresnius rezultatus, prognozėmis testavimo duomenų rinkinyje, supaprastintu būdu apskaičiuoti tikslumo ir subalansuoto F1 įverčiai. Duomenų etiketės buvo priskirtos iš naujo, kategorijas suskirstant į neigiamas, neutralias ir teigiamas. Supaprastinus kategorijų pasiskirstymą, perskaičiuotas modelio tikslumas siekia 83,22%, o subalansuotas F1 įvertis yra 0,8284. Šioje kategorijų sistemoje F1 įvertis neigiamiems komentarams yra 0,8203, neutraliems – 0,5137, o teigiamiems – 0,9117.

4.2. ByT5

Šiame skyriuje pristatomas ByT5 modelis ir šio modelio eksperimentų rezultatai. Trumpai apžvelgiama ByT5 modelio struktūra ir detaliau analizuojami pasirinktos ByT5_{SMALL} versijos parametrų variantai. Pateikiami gauti sentimentų analizės rezultatai..

4.2.1. Modelis ir jo parametrai

Tyrimo metu buvo pasirinkta ByT5 tipo modelio analizė. ByT5 tipo modelis yra paremtas daugiakalbiu mT5 modeliu [58]. Šie modeliai yra pritaikyti dirbti su baitų sekomis, todėl jie gali lengviau apdoroti bet kurios kalbos tekstą bei yra atsparesni natūraliam kalbų triukšmui, kaip gramatinės ir sintaksės klaidos. ByT5 modelių grupė turi keletą skirtingo dydžio modelių. Darbo metu yra analizuojama ir išbandoma ‘maža’ modelio versija. ByT5_{SMALL} modelis yra sudarytas iš 300 milijonų parametrų. Modelio architektūra sudaro $n_{end_blocks} = 12$ koduotojo ir $n_{dec_blocks} = 4$ iškoduotojo bloką, kur $d_{ff} = 3584$, $d_m = 1472$ [58]. Ši ‘maža’ modelio versija yra sudaryta iš didesnio parametrų skaičiaus nei T5_{BASE}.

Siekiant pagerinti sentimentų klasifikavimo uždavinio rezultatus, darbo metu yra koreguojamas ‘ByT5-Lithuanian-gec-100h’¹⁰. Šis modelis buvo papildomai apmokytas lietuvių kalbos gramatikos taisymo uždaviniui [59]. Tai leidžia modeliui geriau gebėti atpažinti ir suprasti lietuvių kalbos

¹⁰ Nuoroda į modelio saugyklą: <https://huggingface.co/LukasStankevicius/ByT5-Lithuanian-gec-100h>

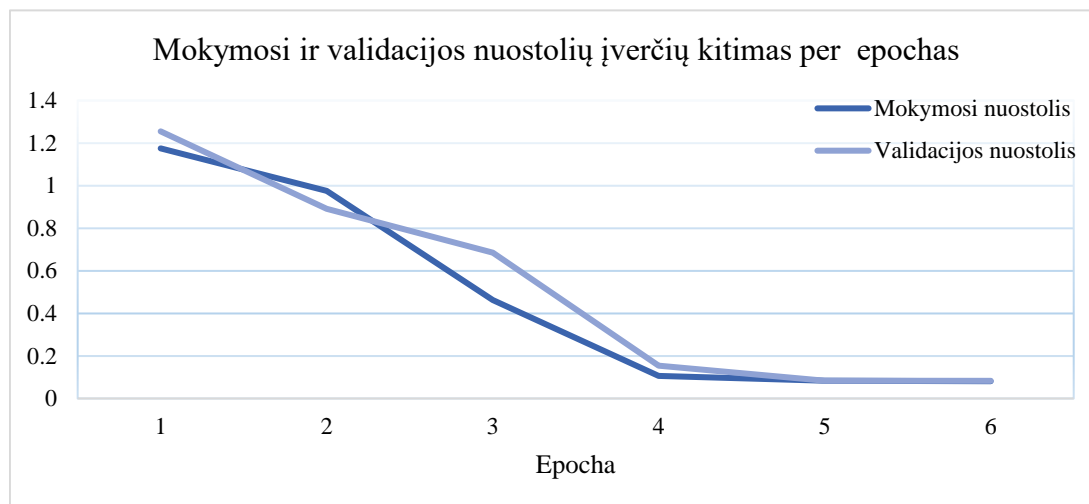
subtilybes. Tikimasi, kad šis modelis gebės lietuviškų sentimentų uždavinio sprendimui pateikti geresnius rezultatus, kadangi literatūroje dažnai minimai aukštesni vienakalbių modelių rezultatai.

4.2.2. Modelio rezultatai

Tyrimo metu buvo tobulinti ir koreguojami du skirtingi ByT5 tipo modeliai pagal turimus du duomenų rinkinius. Šiame skyriuje aptariami geriausi abiejų rinkinių modelių rezultatai.

Šis modelis, pasižymintis pilna transformerio architektūra, pagrįsta koduotuvo–iškoduotuvo schema, sudarytas iš didelio parametų kiekio. Tokio tipo modeliai paprastai reikalauja mažesnio epochų skaičiaus apsimokymui. Ši modelio ypatybė išryškėja apmokant modelį naudojant operatyviai surinktus duomenis. 4.2.1 pav. diagramoje pateikiama modelio, apmokto su pirmuoju duomenų rinkiniu, mokymosi ir validacijos nuostolių įverčių kitimo seka. 1 priede pristatomas antrojo duomenų rinkinio mokymosi nuostolių kitimas. Abiejose diagramose stebimas ryškus mokymosi įverčių sumažėjimas.

Validacijos ir mokymosi įverčių kitimas stabilizuojasi jau ketvirtojoje epochoje. Penktojoje epochoje modelio mokymosi ir validacijos nuostolių įverčiai siekia atitinkamai 0,0838 ir 0,08305. Tai yra žemos reikšmės, kurios idealiu atveju rodo, kad modelis gali atlikti prognozę beveik 100% tikslumu. Tolimesnėse epochose modelis pasiekia puikius mokymosi ir validacijos įverčius, tačiau po šių epochų pastebėtas kitų tikslumo įverčių sumažėjimas. Tai leidžia daryti prielaidą, kad modelis, esant tokio dydžio duomenų rinkiniui, pradeda persimokyti jau penktojoje epochoje.



4.2.1 pav. Pirmojo duomenų rinkinio ByT5 modelio mokymosi nuostolių įverčių kitimo diagrama

Vertinant kitas modelio metrikas, testavimo duomenų rezultatai tarp ketvirtos ir šeštos epochos F1 įvertyje suprastėjo per 0,09 įverčio skalę. Todėl vertinant modelio tikslumą su pirmojo rinkinio duomenimis, buvo pasirinkta naudoti ketvirtos epochos būsenos modelį. Tokiu pačiu analizės principu, su antrąja duomenų imtimi penktos epochos modelis.

Pirmuoju duomenų rinkiniu apmokyto modelio, testavimo duomenų imties prognozės maišaties matrica pateikiama 4.2.1 lentelėje. Lentelė atvaizduoja klasifikavimo rezultatus vertinant modelio spėjimus su tikromis įverčių reikšmėmis visoms penkioms sentimentų kategorijoms. Emocionaliai neigiamų komentarų klasifikavimas pasižymi santykinai mažesniu priskyrimu kitoms kategorijoms. Emocionaliai teigiamų kategorijų prognozės pasiekė aukščiausią tikslumą, tačiau modelis yra linkęs per dažnai prognozuoti emocionaliai teigiamus komentarus. Tačiau neutralių komentarų prognozėse

pastebimas didelis išsiskirstymas, nes modelis beveik vienodu santykiu prognozuoja tiek teigiamus, tiek neigiamus komentarus.

4.2.1 lentelė. ByT5 modelio prognozių maišaties matricos lentelė su pirmuoju duomenų rinkiniu

Prognozės kategorija Tikra kategorija	Emocionaliai neigiamas	Racionaliai neigiamas	Neutralus	Racionaliai teigiamas	Emocionaliai teigiamas
Emocionaliai neigiamas	3753 (81.36%)	0 (0.00%)	594 (12.88%)	32 (0.69%)	195 (4.23%)
Racionaliai neigiamas	792 (38.43%)	138 (6.69%)	737 (35.74%)	60 (2.91%)	115 (5.58%)
Neutralus	629 (20.20%)	3 (0.10%)	1718 (55.16%)	460 (14.76%)	714 (22.96%)
Racionaliai teigiamas	189 (5.32%)	2 (0.06%)	773 (21.75%)	741 (20.84%)	2274 (64.03%)
Emocionaliai teigiamas	155 (1.51%)	1 (0.01%)	371 (3.61%)	342 (3.33%)	9930 (95.53%)

4.2.2 lentelėje pateikiama ByT5 modelio, apmokyto antruoju duomenų rinkiniu, sentimentų kategorijų prognozių maišaties matrica. Šiame modelyje didžiausias kategorijų tikslumas pasiekiamas emocionaliai neigiamų ir teigiamų komentarų kategorijose. Modelis taip pat geriau atpažįsta ir atskiria neigiamo pobūdžio sentimentus. Tačiau neutralių sentimentų prognozės yra gana prastos ir netikslios, modelis linkęs priskirti neutralius komentarus tiek teigiamoms, tiek neigiamoms kategorijoms.

4.2.2 lentelė. ByT5 modelio prognozių maišaties matricos lentelė su antruoju duomenų rinkiniu

Prognozės kategorija Tikra kategorija	Emocionaliai neigiamas	Racionaliai neigiamas	Neutralus	Racionaliai teigiamas	Emocionaliai teigiamas
Emocionaliai neigiamas	3725 (82.80%)	3 (0.07%)	656 (14.58%)	96 (2.13%)	95 (2.11%)
Racionaliai neigiamas	904 (50.55%)	320 (17.89%)	449 (25.11%)	126 (7.04%)	46 (2.57%)
Neutralus	581 (15.60%)	4 (0.11%)	1780 (47.81%)	860 (23.10%)	299 (8.04%)
Racionaliai teigiamas	185 (6.04%)	3 (0.10%)	712 (23.22%)	1940 (63.27%)	1139 (37.13%)
Emocionaliai teigiamas	130 (2.71%)	0 (0.00%)	245 (5.11%)	741 (15.45%)	3458 (72.08%)

Analizuojant abiejų modelių prognozių rezultatus, matoma, kad antruoju duomenų rinkiniu paremtas modelis tiksliau atpažįsta skirtingas kategorijas. Šis modelis yra mažiau linkęs prognozuoti didžiausios tikimybės komentarus, esančius emocionaliai neigiamose ir teigiamose kategorijose, ir sugeba, nors ir netiksliai, atpažinti racionaliais principais pagrįstus komentarus.

4.2.3 lentelėje pateikiami T5, apmokyto su pirmuoju duomenų rinkiniu, apibendrinti prognozių rezultatai trimis sentimentų kategorijoms. Lentelėje matoma, kad modelis 68% tikslumu prognozuoja neigiamus komentarus ir 88% tikslumu atpažįsta teigiamus komentarus.

4.2.3 lentelė. ByT5 geriausio modelio 3 sentimentų kategorijų prognozių maišaties matricos lentelė

Prognozės kategorija \ Tikra kategorija	Neigiamas	Neutralus	Teigiamas
Neigiamas	4683 (68.58%)	1331 (19.50%)	402 (5.88%)
Neutralus	632 (20.30%)	1718 (55.16%)	1174 (37.72%)
Teigiamas	347 (2.78%)	1144 (9.18%)	13287 (88.04%)

4.2.4 lentelėje pateikiami abiejų ByT5 modelių našumo ir tikslumo įverčiai. Bendrai, šio modelio mokymui reikia daug resursų, ir net nedidelis epochų skaičius truko palyginti ilgai. Nors tikslumo įverčių skirtumai nedideli, pastebima, kad abiejuose modelio variantuose testavimo duomenų rinkinio rezultatai yra šiek tiek žemesni, o modelis linkęs prisitaikyti prie mokymo duomenų imties. Be to, pirmasis, didesnės apimties duomenų rinkinys pasiekia geresnius rezultatus.

4.2.4 lentelė. ByT5 modelių našumo ir tikslumo įverčių apžvalga

Modelis	Duomenų rinkinys	Epochų skaičius	Tikslumas, %		Mokymosi trukmė
			Validacijos duomenų rinkinys	Testavimo duomenų rinkinys	
ByT5 _{SMALL}	1	5	65,31	63,26	~ 12 h
ByT5 _{SMALL}	2	6	60,27	59,95	~ 8 h

4.2.5 lentelėje pateikiami ByT5 modelių, naudotų sentimentų analizei, rezultatai, įskaitant atkūrimo ir F1 įverčius. Iš lentelės matyti, kad modelis, mokytas su pirmuoju duomenų rinkiniu, pasiekė geresnius tikslumo rodiklius. Abu modeliai sunkiausiai klasifikuoja racionaliai neigiamus komentarus. Jie gerai atpažįsta emocionališkai neigiamus ir teigiamus komentarus, ypač aukštas F1 įvertis pastebimas dirbant su pirmuoju duomenų rinkiniu. Šiame rinkinyje penktoji kategorija sudaro didžiausią dalį komentarų. Nepaisant to, modelis, mokytas su geriau subalansuotu duomenų rinkiniu, gali tiksliau numatyti sentimentų kategorijas. Pirmasis modelis beveik visiškai nepriskiria racionaliai neigiamų komentarų, o antrasis modelis pasižymi tolygesniu F1 įverčių pasiskirstymu tarp kategorijų.

4.2.5 lentelė. ByT5 modelių sentimentų analizės rezultatai

Modelis	Duomenų rinkinys	Atkūrimo įvertis	F1 įvertis					
			Bendras F1 įvertis	Sentimento kategorijos įvertis				
				1	2	3	4	5
ByT5 _{SMALL}	1	0,6531	0,6032	0,7371	0,0032	0,4424	0,2639	0,8265
ByT5 _{SMALL}	2	0,5955	0,5766	0,7373	0,2541	0,4633	0,5012	0,7295

Galiausiai, ByT5 modeliui, apmokytam su pirmuoju duomenų rinkiniu, tikslumo ir subalansuotas F1 įverčiai buvo apskaičiuoti supaprastintu būdu. Duomenys buvo iš naujo suskirstyti į tris paprastas sentimentų kategorijas, kaip ir distilBERT modelyje. Perskaičiuotas modelio tikslumas siekia 69,14%, o subalansuotas F1 įvertis yra 0,6513. Nors šie rodikliai yra geresni nei penkių kategorijų prognozės, jų skirtumas nėra reikšmingai didesnis.

Apibendrinant, remiantis mokymosi nuostolių kitimo greičiu ir prisitaikymo prie duomenų imties įverčiais, galima teigti, kad šio tipo modeliui, turimi duomenų kiekiai yra per maži norint pasiekti geresnius tikslumo rodiklius.

4.3. GPT-4 Turbo

GPT-4 – tai transformerio architektūra paremtas modelis, sudarytas tik iš išskoduotuvo bloko [60]. GPT-4 buvo pristatytas 2023 metais, OpenAI įmonės. Tai daugiarūšis kalbos modelis, kuris gali apdoroti tekstinę bei vaizdinę informaciją bei generuoti tekstines išvestis. Šis modelis gali būti pritaikytas įvairaus pobūdžio uždaviniuose, kaip tekstų apibendrinimas, vertimas ar pokalbių sistemų kūrimas. GPT-4 sulaukė didelio žmonių susidomėjimo ir tapo vienu žinomiausiu generatyvinio pobūdžio kalbos modeliu. Sprendžiant įvairaus tipo uždavinius, kurie buvo sukurti žmonėms, GPT-4 dažnai sugebėjo gauti geresnius rezultatus nei didžioji dalis žmonių. Pavyzdžiui, simuliuotame advokatų egzamine GPT-4 pasiekė rezultatą, kuris pateko į geriausiai atliktų testų rezultatų 10% [60]. Nepaisant galimybių, modelis nėra visiškai patikimas ir jo taikymuose dažnai susiduriama su modelio atsakymų haliucinacijomis, taip pat modelis yra riboto konteksto ilgio bei nėra permokinamas remiantis savo patirtimi.

4.3.1. Taikomi apribojimai

Šio darbo ribose, siekiant įvertinti modelio gebėjimą prognozuoti lietuviškus sentimentus, modeliui buvo pritaikytas komandų inžinerijos sprendimas. Kadangi modelis yra pritaikytas generuoti tekstą, siekiant analizuoti jo sentimentų analizės kategorizavimo rezultatus – buvo sudaryta tekstinio pobūdžio užduotis. 2 priede pateikiamas užklausų generavimo programinis kodas, kuris kreipiasi į programinę OpenAI sąsają bei gautų rezultatų serializavimas ir apdorojimas. Atliekant užklausas svarbiausi apribojimai:

- nustatoma 0 laipsnių ‚*temperatūra*‘, kuri nurodo modeliui visada grąžinti labiausiai tikėtiną variantą;
- apribojama užklausos apimtis iki 4000 žymų;
- vienos užklausos metu pateikiama ne daugiau nei 30 duomenų įrašų, bet ne daugiau nei 3500 žymų;
- nustatomas 0 baudos taškų pateikimas, kuris leidžia modeliui prognozuoti tuos pačius atsakymus, nepriklausomai nuo jau sugeneruoto teksto. Šis įvertis leidžia modeliui neįtraukti kitų sentimentų užklausos atsakymų į prognozuojamą įvesties rezultatą.

Vertinant GPT-4 sentimentų analizės uždavinį buvo naudojamas antrojo duomenų rinkinio testavimo duomenys. Viso pateikta 636 užklausos ir apdorota 18502 įrašų.

4.3.2. Modelio rezultatai

GPT-4 modelio prognozių antrojo duomenų rinkinio testavimo duomenims painiavos matrica pateikiama 4.3.1 lentelėje. Emocionaliai neigiami ir teigiami komentarai buvo prognozuojami su mažiausia sklaida tarp kitų kategorijų. Modelis beveik nepriskyrė racionaliai neigiamų kategorijų, todėl galima teigti, kad modelis nesugebėjo atskirti emocionaliai neigiamų ir racionaliai neigiamų komentarų. Be to, neutralios kategorijos spėjimai yra gana plačiai pasiskirstę tarp tikrųjų ir prognozuojamų kategorijų.

4.3.1 lentelė. GPT-4 modelio prognozių maišaties matricos lentelė

Prognozės kategorija / Tikra kategorija	Emocionaliai neigiamas	Racionaliai neigiamas	Neutralus	Racionaliai teigiamas	Emocionaliai teigiamas
Emocionaliai neigiamas	4246 (91,12%)	1 (0.02%)	245 (5.32%)	37 (0.80%)	63 (1.37%)
Racionaliai neigiamas	1420 (75.77%)	12 (0.64%)	286 (15.26%)	88 (4.70%)	51 (2.72%)
Neutralus	1302 (32.82%)	4 (0.10%)	1241 (31.30%)	570 (14.37%)	393 (9.89%)
Racionaliai teigiamas	359 (10.16%)	3 (0.08%)	1061 (30.00%)	997 (28.22%)	1554 (43.94%)
Emocionaliai teigiamas	80 (1.95%)	0 (0.00%)	385 (9.38%)	403 (9.81%)	3732 (90.05%)

4.3.2 lentelėje remiantis penkių sentimentų kategorijų klasifikavimo rezultatais, pateikiami GPT-4 modelio apibendrinti prognozių rezultatai trims sentimentų kategorijoms. Lentelėje matoma, kad modelis 87% tikslumu prognozuoja neigiamus komentarus ir 66% tikslumu atpažįsta teigiamus komentarus.

4.3.2 lentelė. GPT-4 modelio 3 sentimentų kategorijų prognozių maišaties matricos lentelė

Prognozės kategorija / Tikra kategorija	Neigiamas	Neutralus	Teigiamas
Neigiamas	5679 (87.71%)	531 (8.20%)	239 (3.69%)
Neutralus	1306 (32.92%)	1241 (31.30%)	963 (24.26%)
Teigiamas	439 (7.77%)	1461 (25.87%)	5286 (66.35%)

4.3.3 lentelėje pateikiami apibendrinti GPT-4 modelio sentimentų analizės rezultatai, įskaitant atkūrimo ir F1 įverčius. Modelio tikslumo įvertis yra 55,18%. Modelio gebėjimas atpažinti skirtingas sentimentų kategorijas įvertintas 0,51 pagal F1 metriką. GPT-4 modelis beveik visiškai nepriskiria racionaliai neigiamų komentarų ir pasižymi reikšmingų klaidų skaičiumi klasifikuojant neutralius ir racionaliai teigiamus komentarus.

4.3.3 lentelė. GPT-4 modelio sentimentų analizės rezultatai

Modelis	Tikslumas, %	Atkūrimo įvertis	F1 įvertis					
			Bendras F1 įvertis	Sentimento kategorijos įvertis				
				1	2	3	4	5
GPT-4	55,18%	0,5518	0,5012	0,7074	0,0132	0,3687	0,3286	0,7183

Galiausiai, GPT-4 modelio prognozės buvo iš naujo suskirstytos į tris paprastas sentimentų kategorijas, kaip ir ankstesniuose modeliuose. Perskaičiavus modelio tikslumą, jis siekia 73,41%, o subalansuotas F1 įvertis yra 0,7013. Šie rodikliai, lyginant tris ir penkias kategorijas, skiriasi nežymiai.

Apibendrinant, GPT-4 modelis pateikė prasčiausius klasifikavimo rezultatus. Net ir supaprastinus sentimentų kategorijas, rezultatų metrikos liko žemiau vidurkio. Kadangi modelis nebuvo specialiai apmokytas šiai užduočiai, tam tikrų klasių atpažinimas gali būti sudėtingesnis.

4.4. Rezultatų palyginimas

Tyrimo metu sentimentų analizės klasifikavimo užduočiai buvo nagrinėjami trijų skirtingų architektūrų modeliai. Du iš jų buvo pritaikyti šiai užduočiai, o GPT-4 modeliui buvo sukurtos specialios užklauskos, pagrįstos užklauskų inžinerija, siekiant klasifikuoti testavimo duomenų rinkinį į penkias kategorijas. Šiandien GPT-4 yra vienas žinomiausių generatyvinių modelių, tačiau jam nebuvo pateikti papildomi mokymo duomenys.

Geriausi rezultatai pateikti 4.4.1 lentelėje. Aukščiausius klasifikavimo rodiklius pasiekė distilBERT_{BASE} modelis. BERT architektūra yra specialiai pritaikyta NLP uždaviniams, susijusiems su teksto žetonais, sekomis, klasifikavimu, apibendrinimu ir atsakymų generavimu. ByT5 modelis taip pat pasiekė gana gerų rezultatų. Skirtingai nuo kitų modelių, jis prieš pritaikant sentimentų klasifikavimui buvo papildomai mokytas lietuviškuose tekstuose.

4.4.1 lentelė. Geriausių sentimentų analizės modelių metrikų įverčiai

Modelis	Duomenų rinkinys	Tikslumas, %	Atkūrimo įvertis	F1 įvertis
distilBERT _{BASE}	1	68,41 %	0,6845	0,6751
ByT5 _{SMALL}	1	63,26 %	0,6531	0,6032
GPT-4	2	55,18 %	0,5518	0,5012

Apskritai, visi modeliai sugeba gana gerai atpažinti emocionaliai teigiamus ir neigiamus komentarus, nes šie dažniausiai pasižymi ryškia emociine raiška ir galimai perteikia daugiau emocionalių išsireiškimų. Lentelės rezultatai rodo, kad distilBERT_{BASE} modelis pasiekė aukščiausią 68,41% tikslumą ir 0,6751 subalansuotą F1 įvertį. Tolesnis modelio pasirinkimas turėtų priklausyti nuo konkrečių poreikių, nes kiekvienas modelis turi savo ypatybių: distilBERT_{BASE} efektyviau atpažįsta įvairesnes sentimentų kategorijas, nors tam tikrose kategorijose pastebima didesnė klaidų sklaida. ByT5 modelis linkęs geriau išskirti emocinėmis savybėmis pagrįstas sentimentų kategorijas.

Galiausiai, vertinant modelius, būtina atsižvelgti į tai, kad lietuvių kalba pasižymi unikalios sintakse ir gramatika. Išankstinio apmokymo metu analizuojami modeliai buvo mokomi su daugiakalbiais duomenų rinkiniais, todėl lietuvių kalba sudaro tik nedidelę mokymo duomenų imties dalį. Kalbos ypatumai bei duomenų rinkinyje esančių atsiliepimų kalbos netaisyklingumas ir buitiškumas daro didelę įtaką sentimentų analizės rezultatams. Surinkti duomenys yra įvairių temų, ir, priešingai nei įprastai atliekamuose tyrimuose, nėra analizuojama specifinė atsiliepimų rūšis. Nors tai padidina modelių pritaikomumą, gali sumažinti rezultatų tikslumą.

Duomenų rinkinyje gausu svetimybų, pastebimas sarkazmas, panieka ir kiti kalbiniai požymiai. Šios ypatybės apsunkina sentimentų analizę, kai nėra papildomos informacijos. Be to, surinktas mokymo duomenų rinkinys yra palyginti mažos apimties transformerių architektūros modeliams. Dėl šių priežasčių kalbos modeliams gali būti itin sudėtinga tinkamai interpretuoti kalbos niuansus ir pasiekti optimalų tikslumą. Derinant modelius su lietuvių kalbos specifika bei plečiant ir gerinant duomenų rinkinių apimtį ir kokybę, galima pagerinti jų veikimo rezultatus. Tai apima papildomą modelių mokymą su didesniu kiekiu lietuviškų duomenų, specifinių kalbos požymių integravimą bei gramatikos ir sintaksės taisyklių pritaikymą modelių veikimui pagerinti.

Išvados

Darbo metu atlikta sentimentų analizės užduoties problemos analizė, suprojektuota ir įgyvendinta prototipinio pobūdžio sentimentų analizės informacinė sistema bei analizuoti, apmokyti ir pritaikyti BERT ir T5 didieji kalbos modeliai sentimentų analizės uždavinio sprendimui.

Išvados:

1. Literatūros apžvalgos metu patvirtinta, kad sentimentų analizė yra aktuali ir sudėtinga natūralios kalbos apdorojimo užduotis. Automatizuoti sentimentų analizės metodai internetinėje erdvėje, padeda gerinti įmonėms įvaizdį, geriau suprasti vartotojų nuomonę, tausoti žmogiškuosius išteklius ir greičiau reaguoti į klientų poreikius.
2. Literatūros apžvalga atskleidė, kad lietuvių kalbos sentimentų analizės tyrimai dažniausiai remiasi tradiciniais klasifikavimo metodais. Tačiau šiuo metu natūralios kalbos apdorojimo srityje geriausius rezultatus pasiekia didieji kalbos modeliai. Apžvalgoje išsamiai aptartos šių modelių architektūros ir pristatytos populiariausios architektūros, taikomos kitų kalbų sentimentų analizėje bei kitose klasifikavimo uždaviniuose.
3. Surinktų duomenų analizė atskleidė, kad laisvai paliktų atsiliepimų kategorijų pasiskirstymas yra netolygus, su didžiausiomis intimis teigiamų ir neigiamų sentimentų kategorijose. Mokymosi proceso metu suvienodinti šį pasiskirstymą su lietuvių kalbos duomenų augmentacija yra sudėtinga dėl reikiamų įrankių trūkumo.
4. Eksperimentų metu buvo koreguojami ir mokomi du transformerių tipo modeliai. Abu modeliai su didesne, bet netolygiai pasiskirsčiusia duomenų intimi parodė geresnius sentimentų klasifikavimo rezultatus, palyginti su mažesne, bet tolygiau pasiskirsčiusia duomenų intimi.
5. Mokymosi epochų metu koreguotų modelių nuostolių pokyčiai parodė, kad modeliai greitai prisitaiko prie surinktos duomenų imties. Tačiau dėl greitos adaptacijos ir mažos duomenų imties modeliai linkę persimokinti. Todėl modelių mokymosi procese buvo taikomas ankstyvas sustabdymas.
6. Analizuojant sentimentų analizės rezultatus, pastebėta, kad visi modeliai sunkiausiai atpažįsta 2 balais įvertintus racionaliai neigiamus atsiliepimus. Be to, modeliai dažnai klaidingai priskiria neutralius atsiliepimus neigiamai kategorijai.
7. Tyrimo metu, naudojant DistilBERT_{BASE} modelį, buvo pasiekti geriausi testavimo rezultatai: subalansuotas F1 įvertis siekė 0,6751, kategorizuojant atsiliepimus į penkias kategorijas, ir 0,8284, vertinant trijų kategorijų sentimentų tikslumą. Apmokinant modelį tik trijų kategorijų prognozavimui, galima tikėtis dar geresnių rezultatų.
8. Geriausio DistilBERT_{BASE} modelio tikslumo ir F1 įverčių skirtumai tarp validacijos ir testavimo rinkinių yra minimalūs, rodantys, kad modelis gerai apibendrina duomenų požymius.
9. ByT5_{SMALL} modelio geriausias pasiektas subalansuotas F1 įvertis, vertinant penkias sentimentų kategorijas, buvo 0,6018. Tyrimo metu modifikuotas ByT5_{SMALL} modelis greitai prisitaikė prie duomenų imties ir sugebėjo gana gerai atpažinti emocionalumu paremtas sentimentų kategorijas.
10. Atliktas eksperimentas parodė, kad GPT-4 modelio, naudoto sentimentų kategorizavimui, rezultatai pasiekė 0,5518 subalansuoto F1 įverčio tikslumą, kas pagerina paprasto spėjimo rezultatus. Tačiau detalūs F1 įverčiai parodo, kad modelis prastai atpažįsta skirtingas sentimentų kategorijas. Iš to galima teigti, kad net ir pažangus kalbos modelis kaip GPT-4 be specifinio pritaikymo gražina prastesnius rezultatus nei užduočiai pritaikytas modelis, turintis mažesnę duomenų bazę ir apmokymo imtį.

Rekomendacijos tolimesniems darbams. Siekiant toliau pagerinti gautus sentimentų analizės klasifikavimo rezultatus, pirmiausia reikėtų didinti duomenų rinkinio apimtį. Surinktas duomenų rinkinys ir rezultatai parodė daug žadančių rezultatų, tačiau norint šią užduotį taikyti praktikoje, reikėtų surinkti didesnę ir tolygiau pasiskirsčiusią duomenų imtį. Be to, dirbant su dideliais kalbos modeliais, labai svarbu turėti iš anksto apmokytus modelius, kurie atitinka užduoties kontekstą. Šiuo metu nėra žinomų vienakalbių lietuvių kalbos didelių kalbos modelių. Vienakalbio modelio sukūrimas ir išankstinis apmokymas yra brangus ir laiko reikalaujantis procesas. Taip pat galima susidurti su nepakankamais mokymo duomenimis dėl ribotų lietuvių kalbos interneto išteklių, palyginti su populiareesnėmis kalbomis, tokiomis kaip anglų ar kinų. Tačiau naujausios tendencijos rodo, kad vienakalbio modelio sukūrimas gali padėti dideliame kalbos modeliui detaliau apibendrinti konkrečią kalbą ir išgauti didesnes kalbos konteksto savybes. Šie metodai galėtų būti toliau tiriami bandant pasiekti dar geresnių klasifikavimo rezultatų.

Literatūros sąrašas

- [1] N. Newman, R. Fletcher and C. Roberts, "Reuters Institute Digital News Report 2022," Oxford, 2022.
- [2] P. Prajna ir R. Anubha, „Opinion mining: a communication tool for career making,“ 2020.
- [3] ReviewTrackers, „Online Reviews Statistics and Trends: 2022 Report by ReviewTrackers,“ 2022.
- [4] EKalba, „Lietuvių kalbos naujažodžių duomenynas,“ 2022.
- [5] A. Balahur Dobrescu, „Methods and resources for sentiment analysis in multilingual documents of different text types,“ Alikantės universitetas, 2011.
- [6] B. Pang ir L. Lee, „Opinion mining and sentiment analysis,“ 2008.
- [7] O. N. a. D. H. P. Balaji, „Levels of sentiment analysis and its challenges: A literature review,“ International Conference on Big Data Analytics and Computational Intelligence, 2017.
- [8] E. M.-C. E. H.-V. a. F. H. C. Zuheros, „Crowd Decision Making: Sparse Representation Guided by Sentiment Analysis for Leveraging the Wisdom of the Crowd,“ IEEE Transactions on Systems, 2022.
- [9] B. Musumali, „An Analysis why customers are so important and how marketers go about in understanding their decisions.,“ 2019.
- [10] R. Y. Kim, „Using Online Reviews for Customer Sentiment Analysis,“ IEEE Engineering Management Review, 2021.
- [11] B. Pang , L. Lee ir S. Vaithyanathan, „Thumbs up? Sentiment Classification using Machine Learning Techniques,“ įtraukta *Empirical Methods in Natural Language Processing*, 2002.
- [12] B. Liu, Sentiment Analysis and Opinion Mining, SpringerLink, 2012, pp. 31-33, 45-60.
- [13] E. Cambria, D. Das, S. Bandyopadhyay ir A. Feraco, A Practical Guide to Sentiment Analysis, Springer, 2017.
- [14] L. C. a. P. Martínez-Barco, „Emotion Detection from text: A Survey,“ 2014.
- [15] A. Collomb, C. Costea, . J. Damien ir O. Hasan, „A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation,“ Lyson Universitetas, Prancūzija, 2013.
- [16] S. D. V. T. N. M. Shelke, „Approaches towards Emotion Extraction from Text,“ pp. 10-14, 2013.
- [17] C. L. J. C. S. Sun, „A review of natural language processing techniques for opinion mining systems,“ 2017.
- [18] P. Thakor ir S. Sasi, „Ontology-based Sentiment Analysis Process for Social Media Content,“ t. 53, pp. 199-207, 2015.
- [19] A. B. Dobrescu, „Extending the EmotiNet Knowledge Base to Improve the Automatic Detection of Implicitly Expressed Emotions from Text,“ European Language Resources Association, 2012.

- [20] E. Cambria, B. Schuller, B. Liu, H. Wang ir C. Havasi, „Statistical Approaches to Concept-Level Sentiment Analysis,“ *IEEE Intelligent Systems*, 2013.
- [21] R. Kohavi, „Glossary of Terms,“ įtraukta *Machine Learning*, Bostonas, Kluwer Academic Publishers, 1998, pp. 271-274.
- [22] A. R. A. T. M. Mohri, *Foundations of Machine Learning*, MIT Press, antras leidimas, 2018.
- [23] R. Gandhi, „Support Vector Machine — Introduction to Machine Learning Algorithms,“ 7 6 2018. [Tinkle]. Nuoroda: <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>. [Kreiptasi 20 11 2022].
- [24] Y. B. A. C. I. Goodfellow, *Deep learning*, Cambridge: MIT Press, 2016, pp. 5-20.
- [25] C. Janiesch, P. Zschech ir K. Heinrich, „Machine learning and deep learning,“ įtraukta *Electron Markets*, Springer Nature Shared, 2021, p. 685–695.
- [26] C. C. O. İrsoy, „Opinion Mining with Deep Recurrent Neural Networks,“ *Association for Computational Linguistics*, 2014.
- [27] I. Sutskever, O. Vinyals ir Q. Le, „Sequence to Sequence Learning with Neural Networks,“ *ArXiv*, 2014.
- [28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser ir I. Polosukhin, „Attention Is All You Need,“ įtraukta *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, 2017.
- [29] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, L. Wei and P. J. Liu, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," in *Journal of Machine Learning Research 21*, Mountain View, 2019.
- [30] S. Chakravarthy, „Tokenization for Natural Language Processing,“ 19 6 2020. [Tinkle]. Nuoroda: <https://towardsdatascience.com/tokenization-for-natural-language-processing-a179a891bad4>. [Kreiptasi 19 11 2022].
- [31] A. Kohli, „Notes on BERT tokenizer and model,“ 25 01 2023. [Tinkle]. Nuoroda: <https://medium.com/@anmolkohli/my-notes-on-bert-tokenizer-and-model-98dc22d0b64>. [Kreiptasi 02 03 2024].
- [32] N. Banik ir H. H. Rahman, „Evaluation of Naive Bayes and Support Vector Machines on Bangla Textual Movie Reviews,“ *International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2018.
- [33] P. Huilgol, „Quick Introduction to Bag-of-Words (BoW) and TF-IDF for Creating Features from Text,“ 28 2 2020. [Tinkle]. Nuoroda: <https://www.analyticsvidhya.com/blog/2020/02/quick-introduction-bag-of-words-bow-tf-idf/>. [Kreiptasi 20 11 2022].
- [34] S. Bagui ir A. Addiga, „Sentiment Analysis on Twitter Data Using Term Frequency-Inverse Document Frequency,“ *Floridos universiteta, Pensakola*, 2022.
- [35] T. Mikolov, K. Chen, G. Corrado ir J. Dean, „Efficient Estimation of Word Representations in Vector Space,“ *ArXiv*, 2013.
- [36] Z. Hlaing, Y. K. Thu, T. Myint, T. Supnithi ir P. Netisopakul, „Supervised Bilingual Word

Embeddings for Low-Resource Language Pairs: Myanmar and Thai,” *įtraukta 16th International Joint Symposium on Artificial Intelligence and Natural Language Processing*, Ayutthaya, Thailand, 2021.

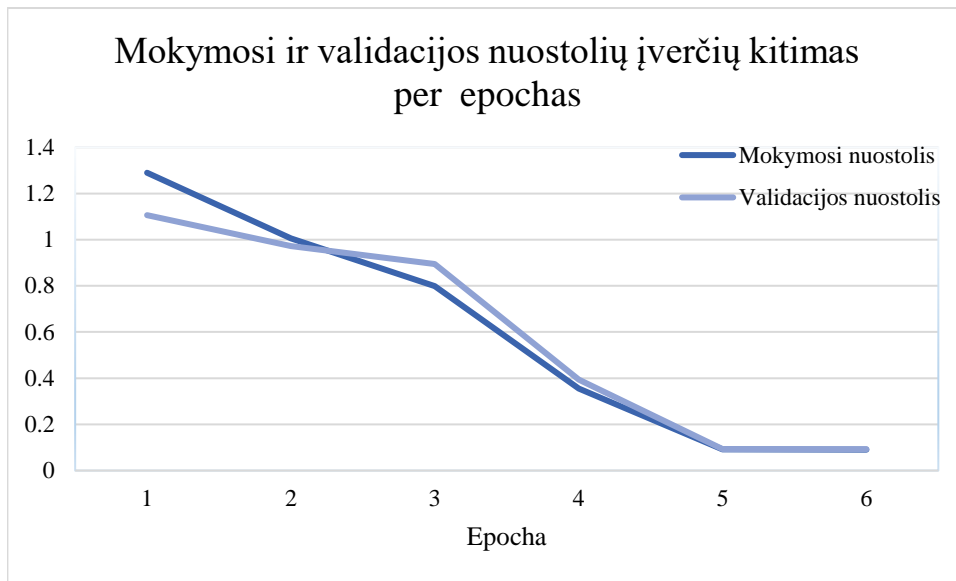
- [37] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee ir L. Zettlemoyer, „Deep contextualized word representations,” 2018.
- [38] A. Neelima ir S. Mehrotra, „A Comprehensive Review on Word Embedding Techniques,” *įtraukta 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCOIS)*, Coimbatore.
- [39] J. Devlin, M.-W. Chan, K. Lee ir K. Toutanova, „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *CoRR*, 2019.
- [40] Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., & Dean, J., „Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation,” *ArXiv*, 2016.
- [41] S. Park, „Sentencepiece: A simple and language-independent subword tokenizer and detokenizer for neural text processing,” 19 05 2023. [Tinkle]. Nuoroda: <https://medium.com/codex/sentencepiece-a-simple-and-language-independent-subword-tokenizer-and-detokenizer-for-neural-text-ffda431e704e>. [Kreiptasi 24 03 2024].
- [42] HuggingFace, „WordPiece tokenization,” HuggingFace, 2023. [Tinkle]. Nuoroda: <https://huggingface.co/learn/nlp-course/en/chapter6/6>. [Kreiptasi 22 03 2024].
- [43] C. Khanna, „WordPiece: Subword-based tokenization algorithm,” *Towards Data Science*, 18 07 2021. [Tinkle]. Nuoroda: <https://towardsdatascience.com/wordpiece-subword-based-tokenization-algorithm-1fbd14394ed7>. [Kreiptasi 24 03 2024].
- [44] S. Khalid, „BERT Explained: A Complete Guide with Theory and Tutorial,” *Medium*, 2019. [Tinkle]. Nuoroda: <https://medium.com/@samia.khalid/bert-explained-a-complete-guide-with-theory-and-tutorial-3ac9ebc8fa7c>. [Kreiptasi 2024].
- [45] N. Reimers, „MLM,” *SBERT.net*, 2024. [Tinkle]. Nuoroda: https://www.sbert.net/examples/unsupervised_learning/MLM/README.html. [Kreiptasi 04 2024].
- [46] c. Sun, X. Qiu, Y. Xu ir X. Huang, „How to Fine-Tune BERT for Text Classification?,” *ArXiv*, 2019.
- [47] A. Merchant, E. Rahimtoroghi, E. Pavlick ir I. Tenney, „What Happens To BERT Embeddings During Fine-tuning?,” *ArXiv*, 2024.
- [48] M. Mosbach, M. Andriushchenko ir D. Klakow, „On The Stability Of Fine-Tuning Bert: Misconceptions, Explanations, And Strong Baselines,” *įtraukta ICLR*, 2021.
- [49] M. Wang, P. Xie, Y. Du ir X. Hu, „T5-Based Model for Abstractive Summarization: A Semi-Supervised Learning Approach with Consistency Loss Functions,” *Applied sciences*, t. 13, 2023.
- [50] T. Kudo ir J. Richardson, „SentencePiece: A simple and language independent subword

tokenizer and detokenizer for Neural Text Processing,” ArXiv, 2018.

- [51] Q. Chen, „T5: a detailed explanation,” Medium, 08 06 2020. [Tinkle]. Nuoroda: <https://medium.com/analytics-vidhya/t5-a-detailed-explanation-a0ac9bc53e51>. [Kreiptasi 24 03 2024].
- [52] D. Shah, „Cross Entropy Loss: Intro, Applications, Code,” V7, 2023. [Tinkle]. Nuoroda: <https://www.v7labs.com/blog/cross-entropy-loss-guide>. [Kreiptasi 10 04 2024].
- [53] A. F. Gad, „Evaluating Deep Learning Models: The Confusion Matrix, Accuracy, Precision, and Recall,” Paperspace, 2019. [Tinkle]. Nuoroda: <https://blog.paperspace.com/deep-learning-metrics-precision-recall-accuracy/>. [Kreiptasi 10 04 2024].
- [54] R. Kundu, „F1 Score in Machine Learning: Intro & Calculation,” V7, 2022. [Tinkle]. Nuoroda: <https://www.v7labs.com/blog/f1-score-guide>. [Kreiptasi 10 04 2024].
- [55] B. Zhu ir W. Pan, „A Text Classification Model Based on BERT and,” įtraukta *4th International Conference on Computers and Artificial Intelligence Technology*, Macao, 2023.
- [56] V. Sanh, L. Debut, J. Chaumond ir T. Wolf, „DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” ArXiv, 2019.
- [57] I. Loshchilov ir F. Hutter, „Decoupled Weight Decay Regularization,” *ICLR*, 2019.
- [58] L. Xue, A. Barua, N. Constant, R. AL-Rfou, S. Narang, M. Kale, A. Roberts ir C. Raffel, „ByT5: Towards a token-free future with pre-trained byte-to-byte models,” ArXiv.
- [59] L. Stankevičius ir M. Lukoševičius, „Towards Lithuanian Grammatical Error Correction,” įtraukta *Computer Science On-line Conference*, 2022.
- [60] OpenAI, „GPT-4 Technical Report,” ArXiv, 2024.

Priedai

1 Priedas. Papildomos ByT5 modelio prognozių rezultatų diagramos



34 pav. Antrojo duomenų rinkinio ByT5 modelio mokymosi nuostolių įverčių kitimo diagrama

2 Priedas. Užklausų į OpenAI programinės sąsają programinis kodas

Toliau pateikiamos pagrindinės Python kalbos programinio kodo iškarpos, skirtos atlikti užklausas į OpenAI programinę sąsają, GPT-4 modelio panaudojimui

```
def RequestOpenApiResponse(messagesToSend):
    messageRequest = [
        {
            "role": "user",
            "content": "Please, can you evaluate the user sentiment of the given review in the
input? I want to receive the sentiment of values:"
            + '"positive" (5 stars), "mostly_positive" (4 stars), "neutral" (3 stars), "negative"
(2 stars), "neutral" (1 star) '
            + " the confidence of that sentiment. If you cannot determine the sentiment, provide
confidence 0 and the sentiment of value"
            + '"unknown". Can you give a JSON response in a format provided between """"
separators?\n'
            + ' """"{"Output":[{"index":1,"sentiment":"positive","confidence":0.75}]}""""\n'
            + f"Input list:\n{json.dumps(messagesToSend, ensure_ascii=False)}",
        }
    ]

    response = openai.chat.completions.create(
        model="gpt-4",
        messages=messageRequest,
        temperature=0,
        max_tokens=4000,
        top_p=0,
        frequency_penalty=0,
        presence_penalty=0,
    )

    return response

<...>

def parse_chat_response(response):
    if not response.choices or not response.choices[0].message:
        raise ValueError("Invalid response format")

    text_content = response.choices[0].message.content # Accessing the content field directly

    if text_content.startswith('""""') and text_content.endswith('""""\n'):
        cleaned_content = text_content[4:-5]
    else:
        cleaned_content = text_content.strip('')

    parsed_data = json.loads(cleaned_content)
    return parsed_data

<...>

for index, row in data_set.iterrows():
    if batchSize == 30:
        print(f"Batch: {batch_no}. Batch size: {batchSize}. Index: {index}")
        messagesToSend.append({"index": index, "review": row["review_text"]})
        response = RequestOpenApiResponse(messagesToSend)
        parsed_response = parse_chat_response(response)
        parsedData = pd.DataFrame(parsed_response["Output"])
        results = pd.concat([results, parsedData])
        messagesToSend = []
```

```
    batchSize = 0
    batch_no += 1
else:
    batchSize += 1
    messagesToSend.append({"index": index, "review": row["review_text"]})
```