

STUDENTŲ ATSLIEPIMŲ VERTINIMAS PASITELKIANT SENTIMENTŲ ANALIZĘ IR BERT TRANSFORMERIŲ MODELIOUS

Ana Usovaitė

Vilniaus kolegija

Anotacija. Straipsnyje nagrinėjamas natūralios kalbos apdorojimo metodų, ypač transformerių architektūros neuroninių tinklų, taikymas studentų atsiliepimų apie dėstomą dalyką analizei. Švietimo institucijoms susiduriant su dideliais tekstinių komentarų kiekiais, tradiciniai analizės metodai tampa neefektyvūs, todėl sentimentų analizė ir teminis modeliavimas suteikia galimybę automatizuotai identifikuoti studentų nuomonės tendencijas, stipriausias ir silpnąsias studijų proceso puses. Tyrime naudojamas daugiakalbis dvikryptis transformacinio kodavimo būdas – (*Bidirectional Encoder Representations from Transformers, BERT*) modelis, gebantis atlikti sentimentų klasifikavimą trijose kategorijose – teigiami, neutralūs ir neigiami. Šis BERT modelis gali būti taikomas net ir tekstams, parašytiems kalbomis, kurių modelis tiesiogiai neapima, pvz., lietuvių kalbai. Analizei taikytas pavyzdinis duomenų rinkinys iš „Kaggle“ šaltinio, kuris sudarytas iš 185 studentų atsiliepimų. Gauti rezultatai rodo aukštą modelio tikslumą atpažįstant teigiamas nuotaikas – su dideliu *tikslumu (precision) (0,954)* ir *įsiminimu (recall) (0,932)*, todėl gaunama aukšta parametro *f1* reikšmė – **0,943**, tačiau mažesnę tikslumą klasifikuojant neigiamas ir neutralias nuomones, kas siejama su duomenų klasių disbalansu. Taip pat atliktas teminis modeliavimas, leidžiantis identifikuoti pagrindines studentų minimas temas, tokias kaip akademinis krūvis, studijų medžiagos kokybė, egzaminų aiškumas ar infrastruktūros trūkumai. Tai leidžia atskleisti, kaip ta pati tema (pvz., egzaminai) vertinama skirtingai – vieni studentai gali teigiamai vertinti egzaminų aiškumą, kiti neigiamai jų sudėtingumą. Tokia metodika atspindi tikrąją nuomonių įvairovę ir leidžia detaliau analizuoti studijų kokybės aspektus. Iš pateiktų rezultatų matyti, kad iš bendrų komentarų skaičiaus 743 komentarų negalima priskirti jokiai temai, nes apklausos dalyviai parašė bendro pobūdžio žodžius arba frazes, tokias kaip *tikslus, geras, patenkinamas* ir t. t. Esant pakankamai dideliame tekstinių atsiliepimų skaičiui, skirstymas pagal atskiras temas leidžia apibendrinti turimą informaciją. Ši metodika suteikia galimybę giliau suprasti studentų patirtis ir priimti duomenimis grįstus sprendimus studijų kokybės gerinimui. Tyrimo rezultatai patvirtina, kad transformerių architektūra yra tinkama ir efektyvi priemonė studentų atsiliepimų analizei, o siūloma metodika gali būti taikoma realių apklausų duomenims, siekiant nuosekliai tobulinti dėstymo procesą.

Reikšminiai žodžiai: transformeriai, BERT, studentų atsiliepimai, sentimentų analizė

Įvadas

Studentų atsiliepimų analizė istoriškai laikoma esminiu dalyku siekiant pagerinti švietimo kokybę. Nuolat besikeičiančioje aukštojo mokslo aplinkoje labai svarbu suprasti studentų požiūrį ir laiku reaguoti į jų problemas, siekiant išlaikyti akademinį pranašumą ir praturtinti bendrą studentų patirtį. Nepaisant to, įprastiems metodams sunku susidoroti su dideliais sugeneruotų tekstinių duomenų kiekiais. Tobulėjant technologijoms, tokioms kaip natūralios kalbos apdorojimas (*Natural Language Processing, NLP*), švietimo įstaigos dabar gali rinkti, apdoroti ir išgauti naudingas žinias iš didelių tekstinių duomenų rinkinių. Natūralios kalbos apdorojimas naudojamas siekiant nestruktūrizuotus tekstinius komentarus paversti struktūrizuotais, pritaikomais duomenimis. Studentų komentarai paprastai suteikia įžvalgių atsiliepimų apie įvairius mokymo įstaigos veiklos aspektus, tokius kaip akademinės programos, dėstytojų veikla, universiteto miestelio infrastruktūra, administracinės paslaugos ir popamokinė veikla. Tačiau rankiniu būdu peržiūrėti šiuos atsiliepimus yra nelengva užduotis, atsižvelgiant į didelį atsakymų skaičių ir įvairovę.

Šio tyrimo **objektas** yra natūralios kalbos apdorojimo taikymas analizuojant studentų atsiliepimus apie dėstomą dalyką.

Šio straipsnio **tikslas** – pasiūlyti metodiką, kuri, naudodama natūralios kalbos apdorojimo (*NLP*) principus (sentimentų analizės ir teminio modeliavimo metodus), analizuotų studentų atsiliepimus apie dėstomą dalyką, tikėdamasi nustatyti stipriausias ir silpniausias veiklos puses bei išsklaidyti abejones.

Tikslui pasiekti išsikelti tokie **uždaviniai**:

1. Išanalizuoti natūralios kalbos apdorojimo, ypač transformerių architektūros neuroninių tinklų taikomojo aspekto studentų atsiliepimų vertinimui mokslinę literatūrą, esamą situaciją ir gerąsias praktikas.
2. Išnagrinėti transformerių architektūros giliojo mokymo pagrindinius aspektus ir galimybes.
3. Sukurti programinę įrangą, kuri, naudodama natūralios kalbos apdorojimo architektūros neuroninių tinklų principus (sentimentų analizės ir teminio modeliavimo metodus), analizuotų studentų atsiliepimus ir įvertintų gautus rezultatus.

Natūralios kalbos apdorojimas apima du pagrindinius komponentus: žmogaus kalbos supratimą ir žmogaus kalbos generavimą. Natūralios kalbos apdorojimas plačiai taikomas įvairiose srityse, tokiose kaip kalbos atpažinimas, mašininis vertimas, atsakymai į klausimus, kalbos sintezė ir sentimentų (nuotaikų) analizė. Nuotaikų analizė turi lemiamą vaidmenį nustatant nuomonių palankumą, negatyvumą ar neutralumą. Tyrėjams gilinantis į natūralios kalbos apdorojimo sritį, visi šie aspektai reikalauja kruopštaus tyrimo. Nuotaikų analizė – tai automatinis duomenų rinkimas iš šaltinio ir jų subjektyvių aspektų nustatymas, siekiant nustatyti, ar vartotojo sukurtas turinys išreiškia palankias, nepalankias ar neutralias nuotaikas (Nandwani, Rupali, 2021; Dang ir kt., 2020). Ši technika apima specializuotas užduotis, tokias kaip nuomonių išskyrimas, aspektų lygmens nuotaikų klasifikavimas ir sakinių lygmens nuotaikų klasifikavimas, dažnai traktuojamas atskirai, naudojant skirtingus modelius, pagrįstus rankinėmis savybėmis arba giliojo mokymosi metodais (Tian ir kt., 2020). Nuotaikų analizė atliekama trimis lygmenimis: dokumento, sakinio ir aspektų lygmenimis. Dokumento lygmens analizė sutelkia dėmesį į bendrą visame dokumente išreikštą nuotaikų supratimą, o sakinio lygmens analizė nagrinėja nuomonių palankumo laipsnį atskiruose sakiniuose. Sakinio lygio analizė yra smulkesnė, nes jos tikslas yra nustatyti sakinių, o ne viso dokumento nuomonę. Aspekto lygmens nuotaikų analizė suskirsto vartotojų nuomones į kategorijas, susijusias su konkrečiais elementais ar atributais. Aspektų lygio nuomonės analizė orientuota į aspektų ar savybių, išreikštų atsiliepimuose, identifikavimą ir vartotojų nuomonių apie šiuos aspektus klasifikavimą (Kastrati ir kt., 2021).

Šio darbo naujumas yra sentimentų analizės taikymas studentų apklausų duomenims apie dalyko dėstymą, papildomai atliekant automatizuotą komentarų suskirstymą į temines kategorijas naudojant transformerių architektūros neuroninius tinklus. Nors sentimentų analizė pati savaime yra plačiai taikoma, jos pritaikymas aukštojo mokslo kokybės vertinimui, studentų atsiliepimų teminei analizei ir automatiniams įžvalgoms apie dėstymo kokybę yra nagrinėtas ribotai. Darbe naudojamas transformerių modelis leidžia ne tik nustatyti emocijų toną, bet ir automatiškai identifikuoti pagrindines temas, kurios studentams kelia pasitenkinimą ar nepasitenkinimą. Tokia metodika suteikia galimybę institucijoms sistemingai analizuoti didelius kiekius atsiliepimų ir priimti duomenimis grįstus sprendimus dėl studijų kokybės gerinimo.

Transformerių architektūros neuroniniai tinklai

Natūralios kalbos apdorojimo transformeriai yra giliojo mokymosi architektūros tipas, kuris stipriai pakeitė šią sritį. Jie leidžia skaitmeniniame pasaulyje efektyviai suprasti ir apdoroti žmogaus kalbą, taikant naujovišką dėmesio (*self-attention*) mechanizmą.

Transformerių modeliai yra daugelio šiuolaikinių NLP programų pagrindas, įskaitant:

- Mašininį vertimą (pvz., „Google“ vertėjas);
- Teksto apibendrinimą;
- Klausimų ir atsakymų sistemas;
- Pokalbių robotus (*chatbots*) ir virtualius asistentus;
- Sentimentų analizę.

Žinomiausi transformerių realizacijos pavyzdžiai yra tokie modeliai kaip dvikryptis transformacinio kodavimo būdas (*Bidirectional Encoder Representations from Transformers, BERT*) ir GPT (*Generative Pre-trained Transformer*). BERT ypač geras teksto supratimui ir yra pažangiausias rezultatais pasižyminti mašininio mokymo metodika, sprendžiant daugumą NLP uždavinių. GPT skirtas teksto generavimui. Šių abiejų modelių pagrindu sudaryta didžioji dalis dabartinių dirbtinio intelekto programų.

Transformeriai sukėlė revoliuciją natūralios kalbos apdorojime, atsisakydami rekurencijos ir pakeisdami ją į dėmesio mechanizmą. Šis mechanizmas leidžia modeliui nustatyti visų žodžių reikšmės svorį sakinyje, nepriklausomai nuo jų padėties, gerai perimdamas kontekstą ir ilgalaikes priklausomybes.

Dėmesio mechanizmo pavadinimas atsirado dėl žmonių gebėjimo selektyviai atkreipti daugiau dėmesio į svarbiausias detales ir ignoruoti mažiau svarbias detales. Žmogus turi visą informaciją, bet sutelkdamas dėmesį tik į pačią svarbiausią informaciją, užtikrina, kad neprarastų jokių reikšmingų detalių, tuo pačiu leisdamas efektyviai naudoti ribotą atmintį ir laiką. Naudojant matematinės išraiškas, apskaičiuojami dėmesio svoriai, kurie atspindi kiekvienos įvesties sekos dalies santykinę svarbą atliekamai užduočiai. Tada jis taiko šiuos dėmesio svorius, kad padidintų (arba sumažintų) kiekvienos įvesties dalies įtaką pagal jos atitinkamą svarbą. Dėmesio modelis – tai yra dirbtinio intelekto modelis, kuriame naudojamas dėmesio mechanizmas, kuris yra apmokytas priskirti tikslus dėmesio svorius taikant prižiūrimą mokymąsi arba savarankišką mokymąsi dideliame pavyzdžių duomenų rinkinyje.

Dėmesio mechanizmą Bahdanau ir kt. (2014) pristatė kaip techniką, skirtą spręsti tuometinių pažangiausių pasikartojančių neuroninių tinklų (*Recurrent neural network, RNN*) modelių, naudojamų mašininiam vertimui, trūkumus. Vėlesni tyrimai integravo dėmesio mechanizmus į konvoliucinius neuroninius tinklus

(*Convolutional neural network, CNN*), naudojamus tokioms užduotims kaip vaizdų subtitravimas ir vizualinis klausimų atsakymas.

Transformatorių modeliai ir juos veikiantys dėmesio mechanizmai pasiekė pažangiausių rezultatų beveik visose gilaus mokymosi srityse. Dėmesio mechanizmų pobūdis suteikia jiems reikšmingų pranašumų, palyginti su konvoliuciniais mechanizmais, naudojamais konvoliuciniuose neuroniniuose tinkluose (CNN) ir pasikartojančiose kilpose, naudojamose pasikartojančiuose neuroniniuose tinkluose (RNN).

Transformeriai pasižymi tokiomis savybėmis:

Lankstumas laike (*Flexibility over time*): RNN apdoroja nuoseklius duomenis iš esmės serializuotai, tai reiškia, kad jie apdoroja kiekvieną sekos laiko žingsnį atskirai tam tikra tvarka. Dėl to RNN sunku atskirti koreliacijas – duomenų mokslo terminologijoje vadinamas priklausomybe – tarp kurių yra daug žingsnių. Dėmesio mechanizmai, priešingai, gali vienu metu nagrinėti visą seką ir priimti sprendimus dėl tvarkos, kuria sutelkti dėmesį į konkrečius žingsnius.

Lankstumas erdvėje (*Flexibility over space*): CNN iš esmės yra lokalūs, naudodami konvoliuciją mažesniems įvesties duomenų pogrupiams apdoroti po vieną dalį. Dėl to CNN sunku atskirti toli viena nuo kitos esančias priklausomybes, pavyzdžiui, koreliacijas tarp žodžių (tekste) arba pikselių (vaizduose), kurie nėra gretimi vienas kitam. Dėmesio mechanizmai neturi šio apribojimo, nes jie apdoroja duomenis visiškai kitaip.

Lygiagretinimas (*Parallelization*): Dėmesio mechanizmų prigimtis reiškia, kad daug skaičiavimo veiksmų atliekama vienu metu, o ne nuosekliai. Tai savo ruožtu leidžia pasiekti aukštą lygiagretaus skaičiavimo laipsnį, pasinaudojant GPU teikiama galia ir greičiu (Bergmann, Stryker, n.d.).

Transformatorių naudojimas švietimo srityje

Apžvalginiam straipsnyje (Pilicita-Garrido, Barra, 2024), nagrinėjama transformerių architektūros modelių taikymo sentimentų analizei švietimo srityje raida ir dabartinė situacija, siekiant nustatyti, kaip šie modeliai padeda suprasti studentų emocijas, nuomones ir patirtis. Autoriai pabrėžia, kad sentimentų analizė tampa vis svarbesnė, nes švietimo institucijos susiduria su dideliais kiekiais tekstinių duomenų – nuo formalių apklausų iki socialinių tinklų įrašų.

Autorių pateiktos išvalgos rodo, kad transformerių modeliai sentimentų analizei švietime pasiekia itin aukštus tikslumo rodiklius, dažnai viršijančius 90 procentų. Jie efektyviai identifikuoja studentų nuotaikas, leidžia aptikti pasitenkinimo ar nepasitenkinimo tendencijas, padeda suprasti mokymosi sunkumus ir dėstymo kokybės problemas. Autoriai išnaginėjo mokslinę literatūrą ir nustatė, kad išryškėjo vienas pastebimas privalumas: nuotaikų analizei naudojamų duomenų objektyvumas, nes yra du pagrindiniai būdai gauti studentų atsiliepimus – tiesioginis ir netiesioginis. Taikant tiesioginį metodą, nuomonės renkamos platinant klausimynus ir vėliau renkant atsakymus. Tačiau šis metodas turi apribojimų, nes neatskleidžia tikrosios studentų patirties ir yra šališkumo galimybė renkant ir vertinant klausimynus. Siekiant įveikti šiuos apribojimus, galima taikyti netiesioginį metodą, kai socialinių tinklų įrašai naudojami kaip informacijos šaltinis renkant studentų nuomones, nes studentai aktyviai naudojami socialine žiniasklaida savo nuomonei išreikšti per įrašus. Šis objektyvumas kyla iš šaltinių platformų, kuriose studentai laisvai reiškia savo nuomonę be išorinio spaudimo, leisdami nustatyti tikrus jausmus, suvokimus ir nuomones. Autoriai teigia, kad transformerių naudojimas ypač naudingas analizuojant netiesioginius duomenis, pavyzdžiui, socialinių tinklų komentarus, kurie atskleidžia autentiškesnes studentų emocijas nei standartinės apklausos. Tai suteikia galimybę švietimo institucijoms gauti objektyvesnį vaizdą apie studentų patirtis ir poreikius (Pilicita-Garrido, Barra, 2024).

Autoriai (Shaik ir kt., 2022) pabrėžia, kad BERT tipo modeliai ypač tinka teksto supratimo užduotims, tokioms kaip klasifikavimas, klausimų–atsakymų sistemos ar vardinių objektų atpažinimas, o GPT modeliai pasižymi išskirtinėmis teksto generavimo savybėmis. Transformerių sėkmę lėmė dėmesio mechanizmas, leidžiantis efektyviai apdoroti ilgus tekstus, išlaikyti kontekstines priklausomybes ir pasiekti didelį lygiagretinimo lygį, kuris neįmanomas naudojant RNN ar CNN. Autoriai išskiria transformerių taikymo iššūkius, tarp kurių – didžiuliai skaičiavimo resursai, reikalingi modelių mokymui, energijos sąnaudos ir aplinkosauginis poveikis. Taip pat aptariama modelių interpretavimo problema: nors transformerių rezultatai yra itin tikslūs, jų sprendimų priėmimo procesas išlieka sunkiai paaiškinamas. Straipsnyje akcentuojama, kad ateities tyrimai turėtų orientuotis į efektyvesnių, mažiau resursų reikalaujančių modelių kūrimą, geresnį modelių paaiškinamumą ir platesnį jų pritaikymą realiose situacijose.

Straipsnyje (Oghu ir kt., 2023) nagrinėjamas sentimentų analizės taikymas studentų atsiliepimų vertinimui, siekiant suprasti jų pasitenkinimą studijų procesu ir identifiuoti pagrindines akademinės aplinkos stiprybes bei trūkumus. Autoriai pabrėžia, kad šiuolaikinėse švietimo institucijose studentų nuomonė tampa svarbiu kokybės rodikliu, tačiau tradiciniai apklausų analizės metodai dažnai neleidžia efektyviai apdoroti didelių tekstinių duomenų kiekių. Todėl sentimentų analizė, paremta mašininio mokymosi algoritmais, suteikia

galimybę automatizuotai ir objektyviai įvertinti studentų emocinį toną bei jų išsakytas problemas. Be to, išskirti metodai, kurie yra daugiausia tiriami siekiant naudoti nuotaukų analizę švietimo srityje. Šie metodai apima prižiūrimą mokymąsi, neprižiūrimą mokymąsi ir leksikonu pagrįstą mokymąsi. Taip pat buvo nustatyti penki mašininio mokymosi algoritmai, taikant prižiūravimo mokymosi metodą: sprendimų medis, paramos vektorių mašina, artimiausio K-kaimyno algoritmas, naiviojo Bajeso algoritmas ir neuroninis tinklas (*Decision Tree, Support Vector Machine, K-Nearest Neighbor, Naive Bayes and Neural Network*).

Tyrimė (Yang ir kt., 2025) siūlomas naujas trumpų tekstų sentimentų analizės modelis, pagrįstas BERT technologija ir dviejų srautų transformerių vartų dėmesio mechanizmu. Šis modelis pirmiausia naudoja BERT ir „Chinese Robustly Optimized BERT Pretraining Approach“ (*Chinese-RoBERTa*). Eksperimentiniai rezultatai parodo modelio pranašumą pasiekiant reikšmingą tikslumo ir f1 rezultato pagerėjimą. Šis rodiklis pasiekė 92,4 %, o tai reiškia vidutinį 8,7 % padidėjimą, palyginti su baziniais modeliais.

Autoriai (Setiawan ir kt., 2023) atliko studentų atsiliepimų analizę. Tyrimė naudojamas BERT modelis, siekiant iš studentų atsiliepimų duomenų išvesti žodžių vektorius. Šiame tyrimė naudojami keli mašininio mokymosi metodai. Nagrinėjamas duomenų rinkinys susideda iš 3 323 studentų vertinimų, o eksperimento rezultatas pagrįstas 80 % mokymo duomenų ir 20 % testavimo duomenų palyginimu. Paramos vektorių mašinos metodas su linijiniu branduoliu pasiekia geriausius rezultatus, kaip rodo 82 % tikslumas ir 90 % f1 vertė.

Straipsnio (Alamoudi ir kt., 2023) tikslas – sukurti duomenų rinkinius arabų kalba, parengtus remiantis Džidos universitete atliktomis studentų pasitenkinimo apklausomis apie studijų dalykus ir dėstytojus. Šiame tyrimė pateikiamas klasikinių mašininio mokymosi modelių vertinimas arabų kalba. Be to, siekiant pagerinti našumą, buvo naudojamas iš anksto apmokytas transformatorius AraBERT, kurio tikslumas siekė 78 %.

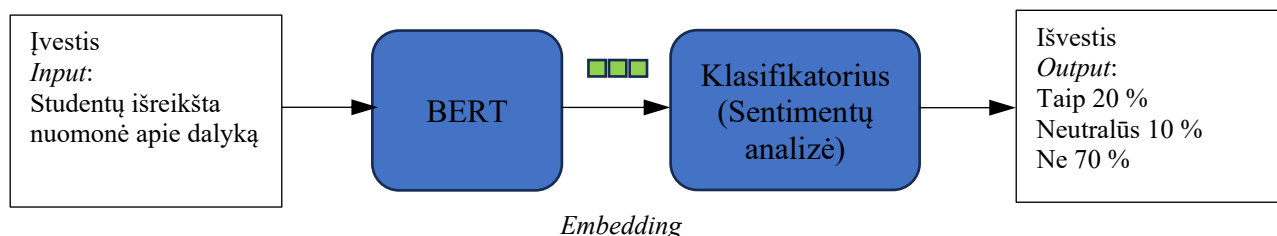
Tyrimo metodologija

A. BERT modelis sentimentų analizei

Sentimentų analizei pasirinktas *nlptown/bert-base-multilingual-uncased-sentiment* modelis, kuris yra daugiakalbė BERT (*Bidirectional Encoder Representations from Transformers*) architektūros versija, specializuota sentimentų analizei. Šis modelis yra iš anksto apmokytas (*pre-trained*) dirbtinio intelekto modelis, skirtas sentimentų (nuotaukų) analizei įvairiomis kalbomis. Tai vienas populiariausių „Hugging Face“ platformoje esančių modelių, skirtų nustatyti, ar tekstas yra teigiamas, neigiamas ar neutralus.

BERT modelis ir jo modifikacijos (pvz., *mBERT, DistilBERT, RoBERTa* ir kt.) sparčiai tobulėja, o jų pritaikymas plečiasi ir į mažesnių resursų kalbas, tokias kaip lietuvių kalba. Tai sudaro prielaidas kurti tikslesnius ir kontekstui jautrius NLP sprendimus įvairioms kalboms.

Modelis apmokytas 6 kalbų (anglų, olandų, vokiečių, prancūzų, ispanų, italų) ir grąžina įvertinimą 1–5 žvaigždučių skalėje. Nors lietuvių kalba nėra tiesiogiai įtraukta į mokymo duomenis, daugiakalbiai modeliai demonstruoja gebėjimą perduoti informaciją tarp kalbų (*cross-lingual transfer*), todėl tinkami analizuoti ir lietuviškus tekstus (Pires ir kt., 2019). Modelio grąžinamas 5 balų įvertinimas konvertuojamas į trijų kategorijų sentimentą: neigiamas (1–2 ★), neutralus (3 ★) ir teigiamas (4–5 ★). Tokia trichotomija yra plačiai taikoma socialinių mokslų tyrimuose, nes leidžia identifikuoti ne tik ekstremalias nuomones, bet ir neutralias ar ambivalentiškas pozicijas (Pang, Lee, 2008). Tekstas apribojamas iki 512 tokenų (*tokens*) – tai BERT architektūros maksimalus įvesties ilgis, užtikrinantis, kad modelis neapdorotų per ilgų sekų, kurios galėtų sumažinti skaičiavimo efektyvumą (Devlin ir kt., 2019). Tokenai yra mažiausi teksto vienetai, kuriuos modelis supranta ir apdoroja. Tai gali būti žodžiai, žodžio dalys, skyrikliai ar net atskiros raidės. Supaprastintas BERT veikimo principas pateiktas 1 pav.



1 pav. BERT modelio veikimo principas

BERT yra iš anksto apmokytų *Transformer Encoder* modelių rinkinys, kuris įveikia ankstesnių modelių apribojimus, tokius kaip paviršutiniškas ELMo dvigubo konteksto sujungimas ir GPT dvikrypčio konteksto trūkumas (Singh, Mahmood, 2021). Gilesnė jo architektūra leidžia tokenams apimti kelis kontekstus,

praturtindama mokymosi aplinką. Šis „Google“ sukurtas modelis pasirodė esąs pranašesnis realiuose taikymuose dėl savo kodavimo įrenginio dizaino ir išsamių mokymų, pagrįstų didžiuliais anglų kalbos „Wikipedia“ ir „Book Corpus“ duomenimis. BERT stengiasi numatyti originalius žetonus, remdamasi aplinkinių žodžių pateiktais kontekstiniais tokenais. Tai leidžia BERT efektyviai užfiksuoti dvikryptes žodžių sakinyje priklausomybes, taip pagerinant kontekstinio žodžių supratimo gebėjimą (Bello ir kt., 2023, Wu ir kt., 2024).

B. Tyrimams panaudoti duomenys

Tyrimui naudojami duomenys iš atviro duomenų šaltinio www.kaggle.com., surinkti studentų apklausos metu. „Kaggle“ yra didžiausia pasaulyje internetinė bendruomenė, skirta duomenų mokslininkams ir mašininio mokymosi (*machine learning*) specialistams. Ji puikiai tinka atlikti mokslinius tyrimus. Adresu <https://www.kaggle.com/datasets/brarajit18/student-feedback-dataset> duomenys pateikti „Excel“ formatu, kuris yra plačiai naudojamas socialinių mokslų tyrimuose dėl savo prieinamumo ir suderinamumo su įvairiais analizės įrankiais. Apklausoje dalyvavo 185 studentai. Duomenys panaudoti kaip pavyzdiniai, vėliau galima siūlomą metodiką taikyti realių apklausų duomenims. Ši duomenų rinkinį sudaro 6 kategorijos, įskaitant dėstymą, kursų turinį, egzaminus, laboratorinius darbus, bibliotekos patalpas ir papildomą veiklą (*teaching, course content, examination, labwork, library facilities, extracurricular*). Kiekvienos kategorijos duomenys apima du stulpelius su skaitiniu ir tekstiniu įvertinimu. Kiekvienas skaitinis įvertinimas gali turėti bet kurią iš trijų žymų: *1 (teigiamas)*, *0 (neutralus)*, ir *-1 (neigiamas)*. Tekstinį įvertinimą sudaro studento parašyta nuomonė. Šis išsamus duomenų rinkinys yra labai svarbus atliekant studentų atsiliepimų nuotaikų analizę, siekiant galutinio tikslo – pagerinti internetinių švietimo pasiūlymų kokybę ir veiksmingumą. Duomenų rinkinio patikimumas ir įvairovė leis suteikti vertingų išvalgų ir išvadų, skatinančių pažangą nuotaikų analizės ir švietimo tyrimų srityje. Vietoje šitų pavyzdinių duomenų galima naudoti realius studentų apklausos duomenis apie dėstomą dalyką. Naudoti pavyzdiniai duomenys specialiai patalpinti platformoje „Kaggle“ tam, kad atlikti skaičiavimus, eksperimentus ar mokslinius tyrimus. Šie duomenys yra patikimi ir tinka preliminariems skaičiavimams, kadangi pagal turinį mažai skirsis nuo realių studentų apklausų apie dėstomą dalyką Lietuvos aukštosiose mokyklose.

Autorė sukūrė programinį kodą „Python“ kalba, kuris sudaro vientisą analizės grandinę, kurioje žingsnis po žingsnio apdorojami studentų komentarai: pirmiausia jie surenkami iš kelių stulpelių ir sujungiami į vieną duomenų rinkinį, tada kiekvienam komentarui priskiriamas sentimentas naudojant daugiakalbį modelį *nlptown/bert-base-multilingual-uncased-sentiment*, o gauti rezultatai papildomi temų modeliavimu su *BERTopic*. Vėliau, remiantis raktažodžiais ir nustatytu sentimentu, komentarams suteikiami žmogui suprantami temų pavadinimai, leidžiantys lengviau interpretuoti bendras tendencijas.

Tyrimo rezultatai ir jų analizė

Tikslumas (Precision) matuoja teisingai identifikuotų teigiamų atvejų (tikrųjų teigiamų) proporciją tarp visų atvejų, kurie buvo prognozuojami kaip teigiami. Jis vertina modelio gebėjimą išvengti neigiamų atvejų klaidingo klasifikavimo kaip teigiamų. Didelis tikslumas rodo mažiau klaidingų teigiamų rezultatų, taip pabrėžiant modelio gebėjimą išlaikyti tikslumą identifikuojant teigiamus nuotaikos atvejus. Tikslumas apskaičiuojamas pagal formulę:

$$\text{Tikslumas} = \text{TP} / (\text{TP} + \text{FP}), \quad (1)$$

kur **TP** – tikrų teigiamų atvejų skaičius; **FP** – neigiamų atvejų, klaidingai klasifikuotų kaip teigiamų, skaičius.

Atgaminimas (Recall) dažnai vadinamas jautrumu, matuoja modelio gebėjimą teisingai identifikuoti visus teigiamus atvejus. Jis matuoja tikrų teigiamų prognozių santykį su bendru faktinių teigiamų atvejų skaičiumi, taip užfiksuodamas gebėjimą išvengti teigiamų atvejų klaidingo klasifikavimo kaip neigiamų. Didelis atgaminimas rodo mažiau klaidingų neigiamų rezultatų, o tai rodo modelio veiksmingumą fiksuojant teigiamus nuotaikos atvejus. Atgaminimas apskaičiuojamas pagal formulę:

$$\text{Atgaminimas} = \text{TP} / (\text{TP} + \text{FN}), \quad (2)$$

kur **TP** – tikrų teigiamų atvejų skaičius; **FP** – teigiamų atvejų, klaidingai klasifikuotų kaip neigiamų, skaičius.

f1-rezultatas (f1-score) yra subalansuotas rodiklis, kuris apima tikslumą ir atgaminimą, kad būtų galima pateikti bendrą modelio veikimo įvertinimą. Tai yra tikslumo ir atgaminimo harmoninis vidurkis, užtikrinantis,

kad būtų įvertinti aukšti abiejų rodiklių rezultatai. *f1-score* yra ypač naudingas dirbant su nesubalansuotais duomenų rinkiniais, kur tikslumas ir atgaminimas gali labai skirtis. Jis apskaičiuojamas pagal formulę:

$$f1\text{-rezultatas} = 2 * ((\text{Tikslumas} * \text{Atgaminimas}) / (\text{Tikslumas} + \text{Atgaminimas})). \quad (3)$$

Parama (Support) reiškia kiekvienos nuomonės klasės faktinių pasikartojimų skaičių duomenų rinkinyje. Ji suteikia vertingą kontekstą interpretuojant *tikslumą*, *atgaminimą* ir *f1-rezultatą*, ypač atvejuose su nesubalansuotu klasių pasiskirstymu.

Tikslumas (Accuracy) atspindi bendrą teisingų prognozių skaičių, padalytą iš bendro prognozių skaičiaus. Nors jis plačiai naudojamas, vertinant nesubalansuotus duomenų rinkinius, jis gali būti nepakankamas, nes aukštą tikslumą galima pasiekti tiesiog prognozuojant daugumos klasę.

Makro vidurkis (Macro Average) apima visų nuotaikos klasių *tikslumo*, *atgaminimo* ir *f1-rezultato* aritmetinio vidurkio skaičiavimą, suteikiant vienodą svorį kiekvienai klasei. Tai ypač naudinga, kai reikia nepriklausomai įvertinti modelio veikimą skirtingose kategorijose.

Svertinis vidurkis (Weighted Average) atsižvelgia į kiekvienos klasės pavyzdžių skaičių, skaičiuojant *tikslumą*, *atgaminimą* ir *f1-rezultatą*. Jis suteikia reprezentatyvesnį vertinimą, kai yra klasės disbalansas, nes atsižvelgia į kiekvienos nuomonės klasės įtaką bendram modelio veikimui.

1 lentelėje pateiktas BERT modelio, skirto studentų internetinių kursų apžvalgų nuotaikų analizei, įvertinimas rodo gerus rezultatus nustatant teigiamas nuotaikas su dideliu *tikslumu (precision) (0,954)* ir *įsiminimu (recall) (0,932)*, todėl gaunama aukšta parametro *f1* reikšmė – 0,943. Tačiau modelis turi tobulėti teisingai nustatydamas neigiamas nuotaikas, ką rodo šiek tiek mažesnė *tikslumo (precision) reikšmė (0,718)*, palyginti su *įsiminimu (recall) (0,503)*. Neutralių nuotaikų atveju modelis pasižymi blogesniu *tikslumu (precision) (0,502)*, tačiau *įsiminimas (recall) (0,713)* rodo kai kurių neutralių apžvalgų neteisingą klasifikavimą. Be to, akivaizdu, kad duomenų rinkinyje yra klasių disbalansas – teigiamų atsiliepimų (792) yra daugiau nei neigiamų (215) ir neutralių (103) atsiliepimų. Šis klasių pasiskirstymo skirtumas gali turėti įtakos modelio veikimui, todėl daugumos klasė (teigiamų) gali gauti geresnius rezultatus, palyginti su mažumos klasėmis (neigiamomis ir neutraliomis). Autorė mano, kad duomenų disbalansas nėra kritinis, nes tikslas yra nustatyti tik bendrą toną ir teigiami atsiliepimai natūraliai dominuoja realiame pasaulyje. Be to, modelis naudojamas tik analizei, o ne mokymui nuo nulio, ir panaudotas iš anksto apmokytas modelis, kuris jau matė subalansuotus duomenis.

1 lentelė. BERT modelio sentimentų analizės rezultatai

	Tikslumas (Precision)	Įsiminimas (Recall)	F1 rezultatas (F1 score)	Parama (Support)
Neigiamas (-1 (1–2 ★))	0,718	0,503	0,592	147
Neutralus (0 (3 ★))	0,502	0,713	0,590	150
Teigiamas (1 (–5 ★))	0,954	0,932	0,943	808
Tikslumas (Accuracy)			0,845	1 105
Makro vidurkis (Macro avg)	0,725	0,716	0,708	1 105
Svertinis vidurkis (Weighted avg)	0,862	0,845	0,848	1 105

Be to, *BERTopic* modulis atliko komentarų paskirstymą pagal temas. Rezultato pavyzdys pateiktas žemiau, 2 lentelėje. Minėtas sąrašas padės nagrinėti teigiamus ir neigiamus aspektus ir teigti pasiūlymus tobulinti konkretaus dalyko dėstymą. Skirstymas į temas vykdomas kaip **dviejų pakopų procesas**, kuriame susilieja dirbtinio intelekto grupavimas ir žmogaus nustatyta logika.

Pirmajame etape *BERTopic* modulis veikia kaip „aklas“ rūšiuotojas: jis nežiūri į žodžių reikšmes, bet skaičiuoja jų statistinį panašumą ir sudeda komentarus į klasterius. Pavyzdžiui, jei dešimtyje komentarų kartojasi žodžiai „skaidrės“, „medžiaga“ ir „pasenusi“, modelis supranta, kad tai yra viena atskira tema, net jei nežino, kas yra skaidrės.

Antrajame etape dirba logikos modulis, kuris tuos klasterių duomenis „įgarsina“. Jis tikrina:

- Turinį: ar tarp pagrindinių grupės žodžių yra specifinių raktažodžių (pvz., *lab*, *exam*, *workload*)?
- Kontekstą: ar BERT sentimento modelis šiuos žodžius įvertino neigiamai (1–2 žvaigždutės) ar teigiamai (4–5 žvaigždutės)?

Galutinis rezultatas yra dinamiškas temų žemėlapis. Pavyzdžiui, jei studentai skundžiasi laboratoriniais darbais, sistema juos priskirs temai „Prastai organizuoti praktiniai užsiėmimai“, o jei giria – „Naudingi

praktiniai užsiėmimai“. Turimi duomenys paverčiami aiškiais įžvalgomis apie tai, kurios studijų sritys veikia puikiai, o kurias reikia tobulinti.

2 lentelė. BERT modelio atlikto komentarų paskirstymo pagal temas rezultatai

Tema	Komentarų skaičius
Bendras komentarų skaičius	1 110
Per didelis akademinis krūvis	61
Kokybiška studijų medžiaga	46
Teigiamas egzaminų aspektas	36
Neskaidri ir sudėtinga vertinimo sistema	34
Aiškūs ir įdomūs dėstymas	31
Nepakankama arba pasenusi studijų medžiaga	30
Naudingi praktiniai užsiėmimai	27
Gera studijų infrastruktūra	26
Valdomas akademinis krūvis	24
Neaiškūs dėstymo stilius	20
Nepakankamos studijų infrastruktūros sąlygos	17
Prastai organizuoti praktiniai užsiėmimai	15
Neaiškios temos	743

Kiekvienam komentarui priskiriamas dinaminis temos pavadinimas, priklausantis tiek nuo *BERTopic* identifikuotos temos (t), tiek nuo individualaus komentaro sentimentu (s). Tai leidžia atskleisti, kaip ta pati tema (pvz., egzaminai) vertinama skirtingai – vieni studentai gali teigiamai vertinti egzaminų aiškumą, kiti neigiamai – jų sudėtingumą. Tokia metodika atspindi tikrąją nuomonių įvairovę ir leidžia detaliau analizuoti studijų kokybės aspektus. Iš pateiktų rezultatų matyti, kad iš bendrų komentarų skaičiaus 743 komentarų negalima priskirti jokiai temai, nes apklausos dalyviai parašė bendro pobūdžio žodžius arba frazes, tokias kaip *tikslus*, *geras*, *patenkinamas* ir t. t. Esant pakankamai dideliame tekstinių atsiliepimų skaičiui, skirstymas pagal atskiras temas leis apibendrinti turimą informaciją.

Skaiciavimuose gauta per daug komentarų, kurie nepriskirti tam tikrai temai, todėl siūloma filtruoti nereikšmingus komentarus, t. y. bandoma nenaudoti tolimesniam apdorojimui komentarų, kurių neįmanoma priskirti konkrečiai temai. Sukurtas komentarų filtro mechanizmas. Jį sudaro keturių sluoksnių komentarų turinio kokybės tikrinimo sistema, kuri analizuoja kiekvieną komentarą ir sprendžia, ar jis yra pakankamai informatyvus, kad būtų vėliau naudojamas *BERTopic* temų moduliui.

Filtas tikrina komentarą iš keturių skirtingų perspektyvų:

- Heuristikos – ar komentaras turi bent minimalią struktūrą;
- „Stopwords“ dominavimo – ar komentaro nesudaro tik tušti žodžiai;
- Entropijos – ar komentaras turi žodžių įvairovę;
- „Embedding norm“ – ar komentaras semantiškai reikšmingas.

Komentaras privalo įveikti visus keturis sluoksnius, kad būtų laikomas informatyviu. Heuristinis filtras – tai pirmasis ir sparčiausiai veikiantis sluoksnis. Jame tikrinama:

- ar komentaras turi bent keturis žodžius;
- ar komentaras nėra tik emocinis žodis;
- ar komentaras nėra šabloninis pagyrimas.

Ši patikra svarbi todėl, kad tokie trumpi komentarai kaip, „*good*“, „*very good*“, „*ok*“, „*nice teacher*“ nesuteikia jokios teminės informacijos. Heuristikos sluoksnyje jie atmetami dar prieš sudėtingesnius skaičiavimus. Šiame žingsnyje tikrinama komentarų struktūra ir pašalinami per trumpi, emociniai komentarai.

Kitas žingsnis yra „Stopwords“ dominavimo patikra. „Stopwords“ – tai žodžiai, kurie neprideda prasmės: *the, is, are, very, really, good, nice, ok, yes, no*. Jei „Stopwords“ sudaro daugiau nei 85 % komentaro – jis atmetamas. Tokie komentarai kaip: „*it is very very good and everything is fine*“, „*the teacher is very nice and very good*“ atrodo ilgi, bet yra semantiškai tušti. Šiame žingsnyje tikrinama komentarų kalbos kokybė ir pašalinami tušti sakiniai.

Trečias žingsnis yra entropijos filtras. Entropijos parametras parodo žodžių įvairovę. Jei komentaras kartoja tuos pačius žodžius, jo entropija maža. Pavyzdžiui, „*good good good good*“ → entropija ≈ 0 , bet „*teacher explains clearly and gives examples*“ → entropijos reikšmė aukšta. Yra nustatoma riba, jei entropijos reikšmė mažiau nei 0,5 ⇒ komentaras atmetamas. Komentarai su maža įvairove yra neinformatyvūs, net jei ilgi. Šiame žingsnyje tikrinama žodžių įvairovė komentaruose ir pašalinami komentarai su pasikartojančiais žodžiais.

Ketvirtas žingsnis yra „Embedding norm“ (*Mashine Learning* filtras). Tai giliausias ir tiksliausias sluoksnis. Jis naudoja „SentenceTransformer embedding“, kad įvertintų semantinį turinį, informacijos tankį ir kontekstinę reikšmę. Kuo šis parametras didesnis, tuo komentaras turtingesnis semantiškai. Tai leidžia atskirti: „good teacher“ → maža „embedding norm“ reikšmė;

„teacher explains concepts clearly and provides examples“ → aukšta „embedding norm“ reikšmė.

Šiame paskutiniame žingsnyje tikrinama mašininio mokymo semantika komentaruose ir pašalinami neinformatyvūs komentarai. Galutiniai filtro veikimo rezultatai pateikti 3 lentelėje.

3 lentelė. BERT modelio atlikto komentarų paskirstymo pagal temas rezultatai taikant komentarų filtravimą

Tema	Komentarų skaičius
Bendras komentarų skaičius	529
Per didelis akademinis krūvis	16
Kokybiška studijų medžiaga	39
Teigiamas egzaminų aspektas	35
Neskaidri ir sudėtinga vertinimo sistema	33
Aiškus ir įdomus dėstymas	41
Nepakankama arba pasenusi studijų medžiaga	26
Naudingi praktiniai užsiėmimai	29
Gera studijų infrastruktūra	25
Valdomas akademinis krūvis	24
Neaiškus dėstyto stilius	51
Nepakankamos studijų infrastruktūros sąlygos	17
Prastai organizuoti praktiniai užsiėmimai	15
Neaiškios temos	178

Iš 3 lentelės duomenų galima matyti, kad atlikus filtravimą gauti geresni rezultatai. Po filtravimo palikti tik aiškūs, ilgesni sakiniai. Filtras palieka tik prasmingus komentarus, leidžia *BERTopic* moduliui geriau suskirstyti komentarus į atskiras temas, bei sumažina triukšmą. Komentarų, priskirtų prie neaiškių temų, skaičius sumažėjo nuo 743 iki 178.

Išvados

Transformeriai pasirodė tinkami studentų apklausų analizei dėl aukšto tikslumo, kontekstinio supratimo ir gebėjimo apdoroti įvairaus stiliaus tekstus. Lyginant su tradiciniais metodais, transformerių modelis pasižymi didesniu našumu ir geresniu semantinių ryšių atpažinimu, todėl yra tinkamas naudoti švietimo kokybės vertinimo procesuose. Atlikta sentimentų analizė parodė, kad transformerių architektūros modelis efektyviai identifikuoja teigiamus studentų atsiliepimus ($f1 = 0,943$), tačiau žemesni neigiamų ($f1 = 0,592$) ir neutralių ($f1 = 0,590$) klasių rezultatai atskleidžia natūralų duomenų disbalanso poveikį. Nepaisant to, metodas leidžia patikimai įvertinti bendrą studentų nuomonės toną ir išskirti pagrindines problemines sritis.

Komentarų suskirstymas į temas parodė, kad studentų atsiliepimai dažniausiai susiję su dėstytojo kompetencija, dėstyto metodu, užduočių aiškumu ir darbo krūviu. Ši teminė analizė leidžia ne tik nustatyti bendrą sentimentą, bet ir tiksliai identifikuoti, kurios studijų proceso dalys kelia daugiausia teigiamų ar neigiamų emocijų. Panaudojus komentarų filtravimą gerėja skirstymų į atskiras temas rezultatai. Atliekant komentarų filtravimą pašalinamas triukšmas, emociniai komentarai, pasikartojantys komentarai, semantiškai tušti komentarai bei paliekami tik tikrai prasmingi sakiniai. Papildomai taikant komentarų filtravimą, *BERTopic* modulis gauna labai švairius duomenis ir skirstymas į temas tampa aiškesnis ir kokybiškesnis.

Gauti rezultatai rodo, kad sentimentų analizės metodika gali būti sėkmingai taikoma ir kitoms studentų apklausų duomenų bazėms, o jos automatizavimas leidžia institucijoms greičiau ir objektyviau vertinti studijų kokybę. Tai ypač aktualu didelėms aukštosios mokykloms, kur rankinis komentarų vertinimas yra neefektyvus. Praktinė rekomendacija švietimo institucijoms: sentimentų analizė gali būti integruota į nuolatinį studijų kokybės stebėsenos procesą, o teminė analizė gali būti naudojama kaip priemonė greitai identifikuoti problemines sritis (pvz., neaiškias užduotis ar perteklinį krūvį) ir priimti duomenimis grįstus sprendimus jų gerinimui.

Ateityje rekomenduojama taikyti naujas BERT modifikacijas, vertinti modelio veikimą skirtingose disciplinose bei integruoti į sentimentų analizę kitus kokybės rodiklius (pvz., pažymius, lankomumą, studentų pasitenkinimo indeksus).

Literatūra

1. Alamoudi, H., Aljojo, N., Munshi, A., & Alghoson, A. (2023). Arabic Sentiment Analysis for Student Evaluation using Machine Learning and the AraBERT Transformer. *Engineering, Technology and Applied Science Research*, 13(5), 11945–11952. <https://doi.org/10.48084/etasr.6347>
2. Bahdanau, D., Cho, K., & Bengio, Y. (2016). Neural Machine Translation by Jointly Learning to Align and Translate. <https://doi.org/10.48550/arXiv.1409.0473>
3. Bello, A., Ng, S-C., & Leung, M-F. (2023). A BERT Framework to Sentiment Analysis of Tweets. *Sensors*, 23(1), 506. <https://doi.org/10.3390/s23010506>
4. Bergmann, D., & Stryker, C. (n.d.). What is an attention mechanism? *IBM*. <https://www.ibm.com/think/topics/attention-mechanism>
5. Dang, N. C., Moreno-Garcia, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, 9(3), 483. <https://doi.org/10.3390/electronics9030483>
6. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
7. Yang, S., Xing, J., Liu, Zh., & Sun, Y. (2025). Short-Text Sentiment Classification Model Based on BERT and Dual-Stream Transformer Gated Attention Mechanism. *Electronics*, 14(19), 3904. <https://doi.org/10.3390/electronics14193904>
8. Kastrati, Z., Dalipi, F., Imran, A., Sh., Nuci, K., P., & Wani, M., A. (2021). Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study. *Applied Sciences*, 11(9), 3986. <https://doi.org/10.3390/app11093986>
9. Nandwani, P., & Rupali, V. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), 81. <https://doi.org/10.1007/s13278-021-00776-6>
10. Oghu, E., Ogbuju, E., Abiodun, T., & Oladipo, F. (2023). A Review of Sentiment Analysis Approaches for Quality Assurance in Teaching and Learning. *Bulletin of Social Informatics Theory and Application*, 6(2), 177–188. <https://doi.org/10.31763/businta.v6i2.581>
11. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/1500000011>
12. Pilicita-Garrido, A., & Barra, E. (2024). Sentiment Analysis With Transformers Applied to Education: Systematic Review. *International Journal of Interactive Multimedia and Artificial Intelligence*, 9(2), 72–83. <https://doi.org/10.9781/ijimai.2025.02.008>
13. Pires, T., Schlinger, E., & Garrette, D. (2019). *How multilingual is multilingual BERT?* <https://doi.org/10.48550/arXiv.1906.01502>
14. Setiawan, H., Fatchah, C., & Saikhu, A. (2023). Multilabel Classification of Student Feedback Data Using BERT and Machine Learning Methods. In *2023 14th International Conference on Information & Communication Technology and System (ICTS)*, 147–152. <https://doi.org/10.1109/ICTS58770.2023.10330849>
15. Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal*, 2, 100003. <https://doi.org/10.1016/j.nlp.2022.100003>
16. Singh, S., Mahmood, A. (2021). The NLP Cookbook: Modern Recipes for Transformer based Deep Learning Architectures. *arXiv preprint arXiv, 2104.10640*. <https://doi.org/10.48550/arXiv.2104.10640>
17. Tian, H., Gao, C, Xiao, X., Liu, H., He, B., Wu, H., Wang, H., & Wu, F. (2020). SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis. *arXiv preprint arXiv*, 2005.05635. <https://doi.org/10.48550/arXiv.2005.05635>
18. Wu, Y., Jin, Zh., Shi, Ch., Liang, P., & Zhan, T. (2024). Research on the Application of Deep Learning-based BERT Model in Sentiment Analysis. *arXiv preprint arXiv*, 2403.08217. <https://doi.org/10.48550/arXiv.2403.08217>

EVALUATION OF STUDENT RESPONSES USING SENTIMENT ANALYSIS AND BERT TRANSFORMER MODELS

Summary

The article examines the application of Natural Language Processing (NLP) techniques, specifically Transformers and BERT models, to analyze student feedback on academic subjects. As educational institutions face vast amounts of textual comments, traditional analysis methods become inefficient; therefore, sentiment analysis and topic modeling offer an automated way to identify trends in student opinions and the strengths and weaknesses of the study process. The study uses a multilingual BERT model for sentiment classification across three categories: positive, neutral, and negative. This BERT model can be applied even to texts written in languages not directly covered by the model, such as Lithuanian. For the analysis, a sample dataset from Kaggle consisting of 185 student reviews was used.

The results indicate high model accuracy in recognizing positive sentiment, with precision (0.954) and recall (0.932), yielding an F1-score of 0.943. However, lower accuracy was observed in classifying negative and neutral opinions, attributed to data imbalance. Topic modeling was also performed, enabling the identification of key themes raised by students, including academic workload, quality of study materials, exam clarity, and infrastructure deficiencies. This reveals how the same topic (e.g., exams) can be perceived differently—some students may positively evaluate exam clarity, while others may negatively assess exam difficulty. Such a methodology reflects the true diversity of opinions and enables a more detailed analysis of study quality.

The results show that, out of the total number of comments, 743 could not be assigned to any specific topic because respondents used general words or phrases such as "accurate," "good," "satisfactory," etc. Given a sufficiently large number of textual feedback entries, categorizing them by individual topics allows for the synthesis of available information. This methodology provides an opportunity for a deeper understanding of student experiences and data-driven decision-making to improve study quality. The research results confirm that the Transformer architecture is a suitable and effective tool for student feedback analysis, and the proposed methodology can be applied to real survey data to consistently enhance teaching quality.

Keywords: transformers, BERT, student feedback, sentiment analysis

Informacija apie autorę

Dr. Ana Usovaitė. Vilniaus kolegijos Elektronikos ir informatikos fakulteto docentė, Vilniaus Tech docentė. Mokslinių tyrimų kryptys: informatikos inžinerija, informacijos sistemos.

El. pašto adresas: a.usovaitė@eif.viko.lt

ORCID: 0009-0001-8380-8165