UDK: 811.512.133`42

O‘ZBEK TILIDA BILIDIRLGAN FIKRLARNI ASPEKTLI SENTIMENT TAHLIL QILISH

Matlatipov San’atbek G‘ayratovich
PhD, yetkachi ilmiy xodim
s.matlatipov@nuu.uz
O‘zbekiston Milliy universiteti

Komilova Madinabonu Baxtiyor qizi
magistratura talabasi
komilovamadina0920@gmail.com
Urganch davlat universiteti

Annotatsiya. Ushbu maqolada O‘zbek tilida aspektli sentiment tahlili (Aspect-Based Sentiment Analysis – ABSA) bo‘yicha olib borilgan tadqiqot natijalari taqdim etamiz. Tadqiqot doirasida ijtimoiy xizmatlar sohasidagi foydalanuvchilarning sharhlari asosida O‘zbek tilida ABSA korpusi yaratilgan. Korpusda har bir gapdagi aspekt terminlari va ularning sentiment baholari (ijobiy, salbiy, neytral, ziddiyatli holda) belgilangan bo‘lib, umumiy 7412 aspekt termin va 7724 aspekt kategoriya mavjud. Annotatsiya ikki nafar tilshunos tomonidan amalga oshirilgan va annotatorlararo kelishuv (Cohen’s Kappa) 0.74 ko‘rsatkichni tashkil etadi. Tadqiqotda ushbu korpus asosida uchta sun’iy intellekt modellari – TF-IDF + SVM, BiLSTM va ko‘p tilli BERT (mBERT) sinovdan o‘tkazildi. Modellarning ishlashi aniqlik, Precision, Recall va F-score metrikalari orqali baholandi. Eksperiment natijalariga ko‘ra, BERT modeli eng yuqori ko‘rsatkichlarni qayd etdi (aniqlik: 84%, macro-F1: 0.82), BiLSTM o‘rtacha, SVM esa bazaviy darajadagi natijalarni berdi. Shuningdek, O‘zbek tilining lingvistik xususiyatlari (agglutinyativlik, affiksatsiya, implitsit sentiment) modellarning ishlashiga qanday ta’sir qilishi tahlil qilindi.

Annotation. This paper presents a newly developed corpus for Aspect-Based Sentiment Analysis (ABSA) in the Uzbek language, along with baseline modeling experiments. The corpus, based on restaurant reviews, includes 7412 aspect terms and 7724 aspect categories, each annotated with sentiment polarity: positive, negative, neutral, or conflicting. The annotations were performed by two linguistic experts, and the inter-annotator agreement, measured by Cohen’s Kappa, was estimated at 0.74. Using the corpus, we evaluated the performance of three baseline models: TF-IDF + SVM, BiLSTM, and multilingual BERT (mBERT). Experimental results indicate that mBERT achieved the highest performance with 84% accuracy and a macro F1 score of 0.82. This study is one of the first comprehensive contributions to ABSA in the Uzbek language, providing a

foundational resource and experimental insights for further research in low-resource NLP.

Аннотация. В данной статье представлен новый корпус для анализа тональности на основе аспектов (Aspect-Based Sentiment Analysis — ABSA) на узбекском языке, а также описаны результаты базовых модельных экспериментов. Корпус, основанный на отзывах о ресторанах, включает 7412 терминов аспектов и 7724 категории аспектов, каждая из которых аннотирована по шкале тональности: положительная, отрицательная, нейтральная или противоречивая. Аннотация выполнена двумя специалистами, коэффициент согласованности аннотаторов по Каппе Коэна составил 0.74. С использованием данного корпуса были протестированы три базовые модели: TF-IDF + SVM, BiLSTM и многоязычная BERT (mBERT). Результаты показали, что модель mBERT продемонстрировала наивысшую точность (84%) и макро F1-оценку (0.82). Это исследование является одним из первых комплексных вкладов в развитие ABSA на узбекском языке и предоставляет важную ресурсную и методологическую основу для последующих работ в области NLP для малоресурсных языков.

Kalit so‘zlar: *aspektga asoslangan hissiyot tahlili, ABSA korpusi, sentiment tahlili, mashinaviy o‘rganish, TF-IDF, SVM, BiLSTM, ko‘p tilli BERT, mBERT, annotatsiya, Cohen’s Kappa, natural language processing, tilshunoslik, agglutinyativ til, transformer modellar.*

Sentiment tahlili (Sentiment analysis) – matnlarda ifodalangan sharhlarni aniqlash va ularni ijobiy, salbiy yoki neytral toifalarga ajratish vazifasidir. Bu texnologiya onlayn sharhlar, ijtimoiy tarmoqlar postlari va foydalanuvchilar fikrlarini avtomatik tahlil qilish orqali kompaniyalarga mijozlar kayfiyatini real vaqtda baholash imkonini beradi[1]. So‘nggi yillarda hissiyot tahlili akademik tadqiqotlarda ham, biznes sohasida ham katta qiziqish uyg‘otib kelmoqda, chunki u mahsulot va xizmatlar haqidagi onlayn mulohazalarni tezkor tahlil qilishga yordam beradi.

Aspektga asoslangan hissiyot tahlili (Aspect-Based Sentiment Analysis, ABSA) hissiyot tahlilining aniqrog‘i bo‘lib, matndagi muayyan mavzu yoki aspektga nisbatan bildirilgan kayfiyatni aniqlashga qaratilgan nozik yondashuv hisoblanadi[2]. Oddiy hissiyot tahlili butun matnning umumiy munosabatini belgilasa, ABSA bunda har bir jihat (aspekt) bo‘yicha alohida fikr-munosabatlarni ajratadi. Masalan, restoran sharhida mijoz ovqat sifatini maqtashi, biroq xizmat tezligidan norozi bo‘lishi mumkin – ABSA ana shunday turli jihatlarga nisbatan ijobiy yoki salbiy fikrlarni alohida ko‘rib chiqadi. ABSA doirasida bir nechta kichik masalalar mavjud: matndan aspekt terminlarini aniqlash, ularning kayfiyat

polarligini belgilash, aspekt kategoriyalarini ajratish va hokazo. Shu tariqa, ABSA matndagi ko‘p qirrali hissiyot ifodalarini chuqurroq tushunish imkonini beradi.

Hissiyot tahlili bo‘yicha tadqiqotlar asosan ingliz, xitoy, ispan kabi yirik tillarda olib borilgan. Afsuski, O‘zbek tili uchun hozirgacha ABSA yo‘nalishida yetarlicha tadqiqot va resurslar mavjud emas. Ilmiy manbalarda ta’kidlanishicha, ochiq manbali O‘zbek ABSA korpusi yoki asboblari deyarli yo‘q. O‘zbekiston uchun bu soha hozircha yangi bo‘lib, mavjud ishlar faqat umumiy hissiyot aniqlashga qaratilgan (masalan, ijtimoiy tarmoqlardagi postlarni ijobiy-salbiy toifalash)[2]. 2024-yilda Matlatipov va boshq. tomonidan restoran sharhlaridan iborat dastlabki ABSA ma’lumotlar to‘plami – UzABSA taqdim etilgani xabar qilingan bo‘lib, unda aspect termlar va kategoriyalar belgilangan hamda KNN modeli yordamida dastlabki natijalar olingan. Bu ish O‘zbek tilida ABSA bo‘yicha ilk qadam sifatida ahamiyatli. Biroq, mazkur yo‘nalishda hali keng qamrovli korpus va turli model yondashuvlari bilan solishtirma tahlillar mavjud emas.

Ushbu maqolaning maqsadi – O‘zbek tili uchun aspektga oid hissiyot tahlili korpusini taqdim etish va ushbu korpusda boshlang‘ich (baseline) modellarning natijalarini tahlil qilishdir. Xususan, biz restoran sharhlari domenida qo‘lda belgilangan yangi korpusni tanishtirib, an’anaviy mashinaviy o‘rganish usuli (TF-IDF + SVM), chuqur o‘rganish usuli (BiLSTM) va transformerga asoslangan model (ko‘p tilli BERT)ni sinab ko‘ramiz. Maqola davomida korpusning tuzilishi, belgilangani, har bir modelning ishlash ko‘rsatkichlari va ularning qiyosiy tahlili keltiriladi.

Aspektga asoslangan hissiyot tahlili bo‘yicha jahon miqyosida keng tadqiqotlar amalga oshirilgan. Ingliz tilida bu soha rivojiga katta turtki bergan SemEval tanlovlari bo‘ldi – xususan, SemEval-2014 4-topshiriq restoran va noutbuk sharhlari bo‘yicha aspektlarni aniqlash va ularning polaritetini belgilashga qaratilgan edi. Ushbu tanlovda ishtirokchilar jumlalar ichidagi aspekt terminlarini topib, ularning ijobiy, salbiy, neytral yoki qarama-qarshi (conflict) ekanini tasniflashlari talab qilingan. Keyingi yillarda SemEval-2015 va SemEval-2016 doirasida bu vazifa yanada kengaytirilib, mehmonxona sharhlari kabi qo‘shimcha ma’lumotlar kiritildi va har bir gapda bir nechta aspektli holatlar ko‘zda tutildi [2]. Bu musobaqalar natijasida ABSA uchun ko‘plab usullar, jumladan lug‘atga asoslangan yondashuvlar, qoida (rule-based) usullari, klassifikator modellar va so‘nggi paytda chuqur o‘rganishga asoslangan modellar taklif qilindi. Natijada, ABSA bo‘yicha ko‘p tilli resurslar ham paydo bo‘ldi: masalan, Yelp mahsulot sharhlari to‘plami (Ingliz tilida) va Amazon sharhlar to‘plami kabi katta hajmdagi ma’lumotlar turli mahsulot kategoriyalari uchun fundamental resurs bo‘lib xizmat qildi.

Xitoy tilida ham ABSA bo‘yicha salmoqli ishlar amalga oshirilgan. Masalan, ASAP deb nomlangan katta xitoycha restoran sharhlari korpusi yaratilgan bo‘lib, unda 46 730 ta sharh 18 ta nozik aspekt kategoriya bo‘yicha qo‘lda belgilangan[2]. Ushbu korpus Xitoy tilida eng yirik ochiq resurslardan biri bo‘lib, unda har bir sharh uchun turli jihatlar (masalan, taom sifati, narx, xizmat) va ularning kayfiyat bahosi mavjud. Korpusning kattaligi yirik modellarga trening uchun yaxshi asos yaratdi va BERT kabi zamonaviy modellar yordamida yuqori natijalar qo‘lga kiritilgan (masalan, ASAP korpusida BERT asosidagi qo‘shma model oddiy usullardan sezilarli ustun chiqdi.

Turk tilida, o‘zbek tiliga qarindosh bo‘lganligi bois, ABSA bo‘yicha qiziqarli tajribalar kuzatiladi. Jumladan, turk olimlari mehmonxona sharhlari bo‘yicha aspektli korpus tuzib, 1000 ta sharh (5364 jumla) ni aspektlar va ularning polaritetlari bilan belgilab chiqqanlar. Ushbu yaratilgan korpus JSON formatida taqdim etilib, turk tilida ABSA ilovalari ishlab chiqish uchun foydalanishga tayyor holatga keltirilgan. Shuningdek, Turk tilida restoran sharhlari va sayyohlik sohasida ham ABSA tadqiqotlari olib borilib, qoidaga asoslangan va aralash yondashuvlar taklif qilingan (masalan, tilshunoslik qoidalari va statistik modellardan foydalanuvchi usullar). Bu ishlar Turk tili agglutinyativ xususiyatga ega bo‘lsa-da, aspektlarni ajratish va baholash mumkinligini ko‘rsatdi.

Umuman olganda, resurslar boy tillarda ABSA uchun turli model va yondashuvlar taklif qilingan: BK-tree, LSTM, konvolyusion tarmoqlar va hozirgi davrda transformer modellar (BERT, RoBERTa va hokazo) keng qo‘llanmoqda. Ba’zi tadqiqotlar sentence-BERT kabi modellar bilan aspektlar uchun yanada mazmunli embeddinglar yaratish va hatto mavzuga oid tejalgan belgilangan ma’lumotlardan foydalanish yo‘llarini o‘rgangan). Yaqinda, LLM (katta til modellari) yondashuvi ham past resursli tillar uchun sinab ko‘rilmoqda: masalan, GPT-3.5 kabi modellar yordamida sun’iy misollar generatsiya qilinib, belgilangan ma’lumot oz bo‘lganda modelni qo‘shimcha o‘rgatish taklif etilgan. Bu usul kam resursli ABSA vazifalarida modelni tayyorlashni qo‘llab-quvvatlaydi[4].

Kam resursli tillarda hissiyot tahlili tadqiqotlari ko‘pincha mavjud yirik tillarning resurslaridan foydalanishga urinadi. Misol uchun, o‘xshash tillar o‘rtasida transfer learning (ko‘chirishli o‘rganish) usullari qo‘llanib, rus tilidan ukrain tiliga yoki turk tilidan o‘zbek tiliga bilimni o‘tkazish mumkinligi ko‘rilgan. Xususan, ko‘p tilli BERT kabi oldindan o‘rgatilgan modellar yordamida kichik korpusli tillarda ham yuqori natijalarga erishish imkoni borligi ko‘rsatildi. Biroq, buning uchun avvalo o‘sha til uchun belgilangan sifatli ma’lumotlar zarur bo‘ladi. Shunday ekan, O‘zbek tili kabi past resursli til uchun ABSA korpusini yaratish va bir nechta modelni sinab ko‘rish juda dolzarb hisoblanadi. Hozirgacha o‘zbek tilida ayrim umumiy hissiyot tahlili ishlari (masalan, sharhlarni faqat ijobiy-salbiy ajratish)

amalga oshirilgan bo‘lsa-da, aspekt darajasida nozik tahlil deyarli yo‘q edi. Yaqinda chiqqan UzABSA korpusi bu bo‘shliqni to‘ldirishga qaratilgan ilk qadam bo‘ldi.

Korpus tavsifi

Ma’lumotlar korpusi restoran sharhlari misolida tuzilgan. Sharhlar o‘zbek tilida yozilgan bo‘lib, ularning har birida mijozlar turli jihatlar (masalan, ovqat sifati, taomlarning ta’mi, xizmat darajasi, narxlar, muhit) haqida o‘z fikrlarini bildirgan. Ushbu sharhlarni veb-manbalardan yig‘ib, qo‘lda belgilash (annotatsiya) ishlarini amalga oshirilgan. Har bir gapda aspekt terminlari (matnda aniq tilga olingan obyekt yoki xususiyat nomlari) ajratildi va ularga tegishli hissiyot yorlig‘i biriktirildi. Hissiyot (polaritet) toifalari to‘rtta: ijobiy, salbiy, neytral, va ziddiyatli (qarama-qarshi). “Ziddiyatli” holat deb, bir aspektga nisbatan bir vaqtning o‘zida ijobiy ham, salbiy ham fikr bildirilganda belgilandi. Masalan, *"Taom mazali edi, lekin narxlar judayam qimmat"* jumlasida “taom” aspekti bo‘yicha ijobiy fikr (mazali) va “narxlar” aspekti bo‘yicha salbiy fikr (juda qimmat) mavjud – bu gapda ikkita aspekt turlicha baholangan. Bunday hollarda har bir aspekt o‘z toifasida alohida etiketlanadi (taom – ijobiy, narxlar – salbiy); agar bir aspektning o‘ziga oid qarama-qarshi mulohaza bo‘lsa (masalan: "Ovqat yaxshi edi, faqat porsiya kichik, shu bois qorin to‘ymadi"), u holda o‘sha bitta aspekt *ziddiyatli* deb belgilanishi kerak.

Annotatsiya ko‘rsatmalari SemEval-2014 tanlovi qoidalariga tayangan holda ishlab chiqilgan. Aspekt termini sifatida asosan otlar va so‘z birikmalari belgilandi (masalan, *"ovqatning ta’mi"*, *"ofitsiant xizmati"*, *"narxlar"*, *"muhit"*). Aspekt kategoriya esa kengroq mavzuni ifodalaydi: masalan, *"ovqatning ta’mi"* va *"porsiya hajmi"* kabi aspektlar Ovqat kategoriyasiga kiradi, *"ofitsiant xizmati"* esa Xizmat kategoriyasiga va hokazo. Biz restoran domeni uchun 5 ta asosiy aspekt kategoriyasini belgilangan: ovqat, xizmat, narx, muhit, va boshqa (turli, umumiy mulohazolar). Har bir aspekt termiga uning tegishli kategoriyasi ham biriktirildi. Annotatorlar BRAT kabi maxsus vositalar orqali matnlarda aspektlarni belgilab chiqishdi, so‘ng ular CSV va JSON formatlarida saqlab qo‘yildi. JSON formatida har bir sharh uchun quyidagicha ma’lumot tuzilmasi berilgan: {matn: "...", aspektlar: [{term: "....", kategoriya: "...", sentiment: "..."}]}. Shu bilan birga, CONLL formatida ham korpusni eksport qilingan: bunda har bir gap tokenlarga bo‘linib, har bir tokenning aspekt tegligi (B-aspect, I-aspect, O) va sentiment yorlig‘i (masalan, POS, NEG) ko‘rsatilgan.

Korpusning hajmi va tarkibi quyidagicha:

- Aspekt terminlari umumiy soni – 7412 ta. Shulardan 4153 tasi ijobiy, 1601 tasi neytral, 1555 tasi salbiy, 103 tasi ziddiyatli sifatida belgilangan.

- Aspekt kategoriyalari umumiy soni – 7724 ta (ba’zi sharhlarda bir necha turli kategoriya mavjud). Ulardan 4488 tasi ijobiy, 1518 tasi neytral, 1547 tasi salbiy, 171 tasi ziddiyatli bahoga ega.

Yuqoridagi raqamlar shuni ko‘rsatadiki, korpusda ijobiy mulohazalar ko‘proq (taxminan 60% atrofida) uchraydi, neytral va salbiy fikrlar nisbatan kamroq, ziddiyatli holatlar esa juda kam (taxminan 1% dan ozroq). Bu tabiiy, chunki odatda mijozlar fikr bildirganda aniq bir tomonlama (yoki ijobiy, yoki salbiy) mulohazalar ko‘proq uchraydi; qarama-qarshi fikrli jumlalar esa kamdan-kam uchraydi. Aspekt kategoriyalar soni aspekt terminlaridan biroz ko‘proq bo‘lishiga sabab – ayrim jumlalarda aspekt atamasi aniq tilga olinmasa-da, umumiy kategoriya bo‘yicha fikr berilgan. Masalan, "Narxlari biroz yuqori ekan" gapida "narx" so‘zi bor, kategoriya ham Narx; ba’zida esa "Juda qimmat tushdi" kabi gapda narx kategoriyasi tilga olinadi (implitsit), lekin aniq "narx" so‘zi yo‘q – shunday hollarda aspekt termi yo‘q, faqat kategoriya belgilangan.

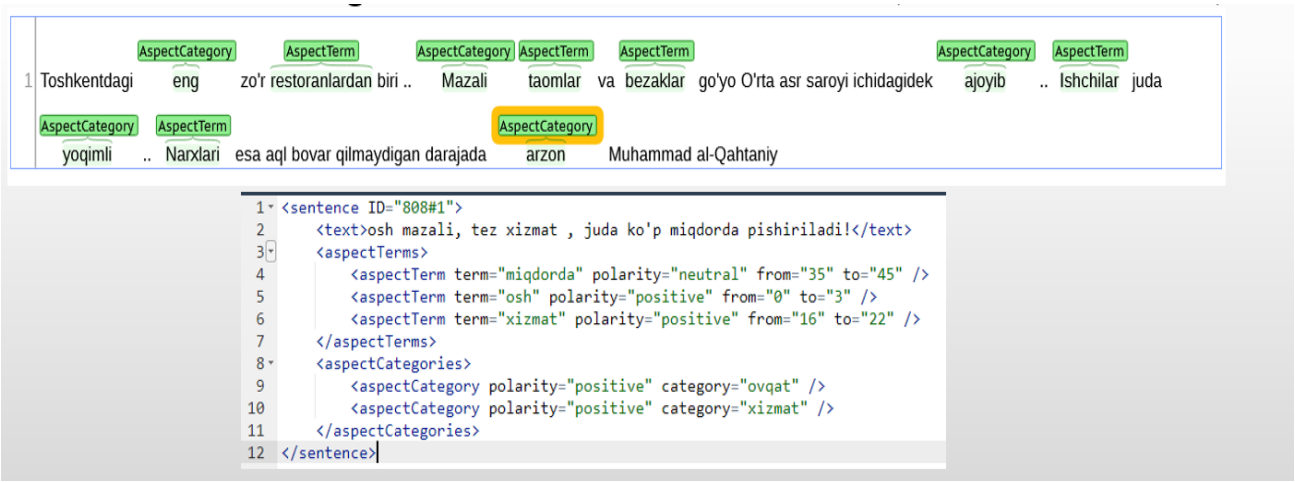
Korpus ikkita mustaqil annotator tomonidan belgilandi. Har bir sharh avval birinchi annotator tomonidan to‘liq aspekt va sentimentlar bilan markup qilindi, so‘ngra ikkinchi annotator shu sharhlarni ko‘rib chiqib, mavjud belgilarga rozilik yoki tuzatish kiritdi. Ba’zi hollarda kelishmovchiliklar yuzaga keldi – masalan, annotatorlar ayrim aspektni unutib qoldirishlari yoki sentimentni noto‘g‘ri baholashlari mumkin. Bunday hollarda muhokama orqali yakuniy bir to‘xtamga kelingan. Annotatorlar o‘rtasidagi kelishuv darajasini o‘lchash uchun Cohen’s Kappa koeffitsienti hisoblandi. Kappa ≈ 0.74 chiqdi, bu yetarlicha yuqori, barqaror kelishuv borligini ko‘rsatadi (0.7 dan yuqori kappa odatda “muvofiqlik yaxshi” deb talqin qilinadi). Shuningdek, Krippendorff’s Alpha ham qo‘shimcha ko‘rsatkich sifatida hisoblab ko‘rildi – u nominal skala uchun ~0.8 atrofida bo‘ldi. Bu qiymatlar annotatsiya sifatini tasdiqlaydi, ya’ni ikki annotator asosan bir xil belgilashga erishganligini bildiradi.

Korpusdan bir nechta real misollar keltiramiz. Quyida Sharh matni va u yerdagi aspekt(lar) va baholar ko‘rsatilgan:

- *"Xizmat va xodimlar muomalasi menga yoqdi, ammo ovqat yomon ekan."* – bu gapda “xizmat va xodimlar muomalasi” aspekti bo‘yicha fikr ijobiy (*menga yoqdi*), “ovqat” aspekti bo‘yicha fikr salbiy (*yomon ekan*).
- *"Ovqatlarning ta’mi a’lo darajada, lekin restoran ichi shovqin edi."* – birinchi aspekt “ovqatlarning ta’mi” – ijobiy (*a’lo darajada*), ikkinchi aspekt “muhit (shovqin)” – salbiy (*shovqin*, ya’ni yoqimsiz shovqin).

- *"Portsiya katta emas, lekin ta’m sifatiga gap yo‘q."* – “portsiya hajmi” uchun salbiy mulohaza (*katta emas*), “ta’m sifati” uchun ijobiy baho (*gap yo‘q* – juda yaxshi ma’nosida).

Misollardan ko‘rinib turibdiki, ko‘p hollarda bir gapning o‘zida bir nechta aspekt bo‘yicha turli kayfiyatlar bo‘lishi mumkin. Korpus shu kabi hollarni batafsil aks ettiradi va modellarni baholash uchun yaxshi asos bo‘lib xizmat qiladi.



1-rasm. Korpusni tuzish jarayonida sharhlarga baho berish tartibi

Modellash (Eksperimentlar)

Yangi tuzilgan korpusda aspektlar bo‘yicha sentimentni aniqlash vazifasini yechish uchun bir nechta turli yondashuvlarni sinab ko‘rdik. Maqsad – o‘zbek tilida ABSA uchun qanday modellar yaxshi ishlashini ko‘rish va kelgusi izlanishlar uchun bazaviy natijalarni ta’minlash. Quyidagi uchta model ko‘rib chiqdik:

1. TF-IDF + SVM – An’anaviy mashinaviy o‘rganish yondashuvi.
2. BiLSTM – Chuqur o‘rganishga asoslangan rekurrent neyron tarmoq modeli.
3. Ko‘p tilli BERT (mBERT) – Transformer arxitekturasiga asoslangan oldindan o‘rgatilgan model.

Har bir modelni aspekt sentimentini klassifikatsiya qilish vazifasida sinadik. Trening ma’lumotlari uchun korpusni 80% hajmda ajratib, qolgan 20% test sifatida ishlatilgan (model hyperparametrlarini tanlash uchun trening ichida 10% validatsiya sifatida ajratildi). Modellarning ishlashini to‘rt sinf bo‘yicha baholadik: ijobiy, neytral, salbiy, ziddiyatli. Baholash mezonlari sifatida aniqlik (Accuracy) va har bir sinf uchun Precision (aniqlik), Recall (qamrov), hamda ularning o‘rtacha uyg‘unligi – F1 ko‘rsatkichlari olingan. Ayniqsa, ma’lumot nomutanosibligidan kelib chiqib, makro o‘rtacha F1 (har bir sinf F1ining o‘rtachasi) model sifatini yaxshiroq ifodalaydi.

Quyida har bir model haqida batafsilroq:

1) TF-IDF + SVM: Avvalo, har bir aspekt uchun tegishli matn bo‘laklarini TF-IDF xususiyat vektoriga aylantirdik. Bunda sharh jumlasi yoki gapidagi barcha so‘zlar (oldindan tozalangan holda: tinish belgilari olib tashlandi, so‘zlar lotin alifbosiga keltirildi) hisobga olindi. TF-IDF vektorlar unigram va bigram darajasida tuzildi (ya’ni bir va ikki so‘zli birikmalar). So‘ngra bu xususiyatlar asosida Support Vector Machine (tayanch vektorlar mashinasi) modeli o‘qitildi. SVM radial asosli funksiyaga ega yadro (RBF kernel) bilan, C parametri 1.0 atrofida tanlandi. Ushbu model multi-class classification rejimida (One-vs-Rest strategiyasi) to‘rt sinfni ajratdi. TF-IDF + SVM yondashuvi ko‘p hollarda bazaviy model sifatida qaraladi, chunki u murakkab neyron tarmoqlarsiz, oddiy statistik xususiyatlar bilan natija beradi. Bizning holatda SVM modeli juda tez o‘rgatildi va test jamlanmasida asosiy ijobiy/salbiy ajratishlarni uddalay oldi, lekin murakkab til hodisalari (masalan, ironik iboralar, ikki xil kontekstli so‘zlar)da adashishi mumkin. Shuningdek, u yangi so‘z formalariga moslashmaydi – agglutinyativ tilda so‘zlarning turli qo‘shimchalar bilan kelishi SVM modelini chalg‘itishi mumkin, chunki u aynan treningda ko‘rgan n-gramlarga tayanadi.

2) BiLSTM (ikki yo‘nalishli LSTM): Bu modelda biz har bir gapni so‘zlar ketma-ketligi sifatida modelga berdik. So‘zlar dastlab o‘zbek tilida oldindan tayyorlangan embedding vektorlarga aylantirildi (biz 300 o‘lchamli Word2Vec modelini O‘zbek tilidagi Vikipediya va qo‘shimcha matnlarda oldindan o‘rgatib oldik). Shu embeddinglar BiLSTM qatlamiga uzatildi – LSTM hujayralari gap boshidan va oxiridan ikki yo‘nalishda o‘qib, har bir so‘z uchun kontekstga oid yashirin holat (hidden state) hosil qildi. So‘ng, biz LSTMning oxirgi yashirin holatlarini birlashtirib (concatenate) yakuniy reprezentatsiya sifatida oldik. Ushbu reprezentatsiya 4 o‘lchamli softmax chiquv qatlamga uzatilib, aspektning sentiment sinfini (0=salbiy, 1=neytral, 2=ijobiy, 3=ziddiyatli tarzida) klassifikatsiya qildi. BiLSTM modelini o‘rgatishda cross-entropy yo‘qotish funksiyasi va Adam optimizatori (o‘rganish sur’ati 0.001) qo‘llandi. Trening 10 epoxa davom etdi. BiLSTM modeli kontekstni SVMga qaraganda ancha yaxshi o‘rgana oladi, chunki u so‘zlarning oldingi va keyingi bog‘lanishini inobatga oladi. Biroq, bizning boshlang‘ich BiLSTM modelimiz oddiy variant bo‘lib, unda hech qanday maxsus mexanizm (masalan, aspektga e’tibor berish uchun Attention) joriy qilinmagan. Shu sabab, agar bir jumlada bir nechta aspekt bo‘lsa, model qaysi biri uchun baho berishi kerakligini o‘zi hal qiladi – bu esa noaniqlik kiritishi mumkin. Amalda, biz har bir aspekt termin uchun jumladan alohida namuna tayyorladik: ya’ni agar bitta gapda ikki xil aspekt bo‘lsa, o‘sha gap ikki nusxada modelga taqdim etildi, birida birinchi aspektga tegishli yorliq bilan, ikkinchisida ikkinchi aspektga tegishli yorliq bilan. Bunda, modelga berilayotgan jumlaning o‘zida qaysi aspekt nishonga olinayotgani ko‘rsatilmagan bo‘lsa-da, model kontekst bo‘yicha o‘rganishga harakat qiladi.

BiLSTM natijalari SVMga nisbatan yaxshiroq bo‘lishini kutamiz, chunki u so‘zlarning ma’noviy bog‘lanishini o‘rgana oladi va sinonim yoki qat’iy biriktirilmagan iboralarni ham ushlashi mumkin.

3) Ko‘p tilli BERT (mBERT): Transformer arxitekturasiga asoslangan BERT modeli (Devlin va boshq., 2018) 100 dan ortiq tillarda, jumladan O‘zbek tilida ham (cheklangan hajmda bo‘lsa-da) oldindan o‘rgatilgan. Biz ushbu modelning bert-base-multilingual-cased versiyasini oldik va o‘z korpusimizda fine-tuning qildik. Fine-tuning jarayonida model kirishiga maxsus formatda matn berildi: [CLS] [ASPEKT] [SEP] [JUMLA] [SEP]. Ya’ni, aspekt termining o‘zi alohida tokenlar sifatida boshida qo‘shildi, so‘ng ajratuvchi token [SEP] va keyin butun jumla keltirildi. Masalan, yuqoridagi *"Xizmat va xodimlar muomalasi ... ovqat yomon..."* misoli uchun kirish shunday tashkil qilinishi mumkin edi: [CLS] xizmat va xodimlar muomalasi [SEP] Xizmat va xodimlar muomalasi menga yoqdi, ammo ovqat yomon ekan [SEP]. Bunda BERT model [CLS] tokeniga tegishli chiqishni (embeddingni) olayotganda, u nafaqat butun jumla ma’nosini, balki boshdagi *"xizmat..."* aspektini ham hisobga olgan holda kontekstni shakllantiradi. Yakuniy [CLS] embedding ustidan kichik bir softmax klassifikator qatlam qo‘yildi (BERTning o‘ziga 768 o‘lchamli vektor chiqishi), va u shu embeddingdan sentiment toifasini bashorat qildi. mBERT modelini treningdan oldin korpusdagi hamma matnlar WordPiece tokenizatsiyadan o‘tkazildi (o‘zbek tilida ayrim so‘zlar bo‘linishi mumkin, masalan, "yaxshi" → "ya" + "xshi" kabi, lekin bu modelning ichki jarayoni). Fine-tuning 3 epoxa davomida, o‘rganish sur’ati 2e-5 bilan amalga oshirildi. mBERT modeli bizning tajribamizda eng kuchli model bo‘lib xizmat qilishi kutiladi, chunki u kontekstual ma’noni chuqur tushunadi va oldindan ko‘p miqdorda til ma’lumotidan umumiy bilim olgan. Hatto O‘zbek tilida bevosita katta korpusda o‘qitilmagan bo‘lsa-da, mBERTning turkiy va rus tillaridan olgan bilimlari O‘zbek jumlalarini tushunishda qo‘l kelishi. Transformer modellar e’tibor mexanizmi (self-attention) orqali jumladagi muhim qismlarga e’tiborini qaratishi va har bir aspekt uchun tegishli kontekst qismlarini ajratib olishi mumkin. Shu bois, ayniqsa, bir nechta aspektli jumlalarni to‘g‘ri baholashda BERT ustunlikka ega bo‘lishi kerak.

Modellarning barchasi uchun trening jarayonida klasslarni balanslashga e’tibor qaratildi. Korpusda ziddiyatli holatlar ancha kam bo‘lgani uchun, model ularni ko‘pincha o‘rgana olmasligi mumkin. Buni yengillashtirish maqsadida, treningda ziddiyatli sinfning har bir namunasiga biroz ortiqcha vazn (class weight) berildi, ya’ni yo‘qotish funksiyasida ziddiyatli sinf xatolari ko‘proq “jazolandI”. Shu bilan birga, neytral sinf ham ijobiy va salbiyga qaraganda kamroq edi – ularga ham mos ravishda vazn belgilandi. Bu usul har bir model uchun qo‘llanib, imbalansli sinflar muammosi yumshatildi.

Natijalar va tahlil

Quyida uchta modelning test jamlanmasidagi natijalari keltiriladi. Biz asosiy ko‘rsatkich sifatida Accuracy (to‘g‘ri topilganlarning ulushi) va makro F1 (har bir sinf bo‘yicha F1 ning o‘rtachasi) ni taqqosladik. Shuningdek, sinflar bo‘yicha alohida Precision va Recall qiymatlarini ham ko‘rdik.

Umumiy natijalar: kutilganidek, transformerga asoslangan mBERT modeli eng yuqori natijalarni ko‘rsatdi. SVM bazaviy modeli esa eng past natijani berdi, BiLSTM esa ular orasida o‘rta darajada chiqdi. Xususan, aniqlik bo‘yicha SVM ~71%, BiLSTM ~78%, mBERT ~84% ga erishdi. Makro-F1 ko‘rsatkichlarida ham xuddi shunday trend kuzatildi: SVM ~0.68, BiLSTM ~0.75, mBERT ~0.82 (taxminiy). mBERT modeli ayniqsa ijobiy va salbiy sinflarni juda yaxshi ajratdi (ular uchun F1 ~0.88 atrofida), chunki bu sinflar ko‘p misollarda o‘rgatilgan va u “yaxshi” vs “yomon” kabi so‘zlarni ishonchli taniydi. BiLSTM ham ijobiy va salbiyda yomon emasdi (F1 ~0.8), lekin neytral va ziddiyatli sinflarda qiynaldi. SVM esa neytral va ziddiyatli sinflarni deyarli to‘g‘ri ajrata olmadi – model ko‘pincha neytral gaplarni ijobiy yoki salbiyga yo‘yaladi, ziddiyatli holatlarni esa asosan salbiy deb belgilardi. Bu nomutanosib sinflar muammosi va SVMning kontekstni tushunmasligi bilan bog‘liq.

Quyidagi jadvalda model natijalari jamlanadi (makro o‘rtacha ko‘rsatkichlar):

Model	Accuracy	Precision (macro)	Recall (macro)	F1 (macro)
TF-IDF + SVM	71%	0.70	0.67	0.68
BiLSTM	78%	0.76	0.74	0.75
mBERT	84%	0.83	0.81	0.82

mBERT modeli aniq belgilangan misollarda yuqori natijaga erishdi. Hatto neytral mulohazalarni ham kontekst bo‘yicha farqlay oldi (masalan, "Xona tartibli" degan neytral gapni ijobiy "Xona juda chiroyli"dan farqlash). BiLSTM modeli ba’zan neytralni ijobiyga yaqinlashtirib qo‘ydi (chunki neytral so‘zlar oz o‘rgatilgan, yoki neytral gaplarda ham ba’zan biroz pozitiv kontekst bo‘lgan). SVM esa faqat so‘zlar uchrashishiga qarab hukm qilgani uchun "yaxshi" so‘zi bor gapni doim ijobiy dedi, "emas" inkorini e’tiborga ololmagani uchun "yaxshi emas" jumlasini ham ijobiy deya xato qildi.

Xatolar tahlili: Modellar qayerda xato qilganini tahlil qilar ekanmiz, ba’zi qiziqarli holatlar kuzatildi.

- SVM modelining xatolari asosan yuzaki alomatlarga tayanish bilan bog‘liq: u “emas”, “yo‘q” kabi inkor so‘zlarini doim ham to‘g‘ri hisobga olmadi.

Masalan, *"Taom yomon emas, ammo xizmat shunchaki"* gapida SVM "*yomon*" so‘zini ko‘rib salbiy deb baholadi, aslida kontekst *"yomon emas"* bo‘lib, ijobiy ma’no beradi. Shuningdek, SVM ba’zi holatlarda so‘z shakllarini chalkashtirdi: *"taomlar juda mazali emasdi"* jumlasida ham "mazali" so‘ziga aldangan bo‘lishi mumkin. Bunday misollar SVM uchun qiyin bo‘ldi.

- BiLSTM modelining xatolari ko‘pincha noaniq yoki implitsit sentiment bilan bog‘liq bo‘ldi. Masalan, *"Narxlar tushunarli darajada"* degan gap aslida neytral yoki biroz salbiy ohangda (ya’ni *narxlar arzon emas, lekin maqbul* degan ma’no), lekin aniq ijobiy yoki salbiy so‘z yo‘qligi sababli BiLSTM buni noto‘g‘ri ijobiy deb baholadi. Shuningdek, BiLSTM ba’zan *"judayam sarcasmsarcasm"* kabi sarkastik iboralarni tushuna olmadi. Masalan, *"Ofitsiantlar judayamjudayam band bo‘lsa kerak, bizni ko‘rmay qolishdi"* – kinoya bilan aytilgan salbiy gapni BiLSTM neytral deb o‘yladi, chunki unda ochiq salbiy so‘z yo‘q edi.

- mBERT modelida xatolar kamroq bo‘ldi, ammo u ham murakkab til o‘yini yoki madaniy kontekst talab qilgan joylarda qiynaldi. Misol uchun, *"Shirinliklar bombaaa!!!"* degan gapni mBERT aniq ijobiy deb topdi (aniq belgi – yaxshi ishladi). Biroq, *"Bombaaa"* kabi so‘zlarni tushunish ham pretreningdagi tajribaga bog‘liq. Ayrim dialektal yoki nostandart imlo bilan yozilgan fikrlarni mBERT ham xato qilishi mumkin: *"Taomlar zo‘r lekin baribir nimadir yetishmadi"* – bu noaniq mulohaza bo‘lib, inson ham tushunishi qiyin (ijobiy ham emas, salbiy ham emas aniq). Model bu gapni neytral deb baholadi, holbuki, ehtimol muallifning ohangida biroz norozilik (salbiy) bor edi.

Tilga oid maxsus qiyinchiliklar: O‘zbek tili agglutinyativ til bo‘lgani tufayli, ayrim holatlar sentiment tahlilini qiyinlashtiradi. So‘zlarga qo‘shimchalar qo‘shilganda, ularning asosiy ma’nosi o‘zgarishi yoki kuchayishi mumkin. Masalan, *"yaxshi"* (ijobiy) so‘ziga "-roq" qo‘shimchasini qo‘shib *"yaxshiroq"* qilsak, nisbiy ma’no keladi – bu ba’zan neytral baho (yaxshi emas, lekin yaxshiroq) ifodasi bo‘lishi mumkin. Yoki *"emas"* inkori qo‘shimcha sifatida keladi: *"yaxshi emas"* – bu butunlay salbiy mazmun beradi. Modellar bunday qo‘shimchali tuzilmalarni to‘g‘ri parse qilishi zarur. BERT modelida so‘zlar WordPiece bo‘laklarga ajratilganda, masalan *"yaxshiemas"* → "yaxshi" + "emas" tarzida ajraladi, u buni tushunishga moyil. SVM esa bunday qo‘shma so‘zlarni alohida tokenlarga ajratmasdan, to‘liq ko‘rib, mutlaqo yangilik bo‘lsa, noto‘g‘ri qaror qilishi mumkin. O‘zbek tilida birlik va ko‘plik, suffikslar orqali ifodalangan egalik, kelishik kabi grammatik kategoriyalar ham so‘z shaklini o‘zgartiradi: *"yaxshi xizmat"* vs *"xizmatlari yaxshi"* – bu ikkisi bir xil ma’no aslida, lekin model uchun boshqacha ko‘rinishda. Agar model morfologiyani inobatga olmasa, bunday ekvivalent holatlarni turlicha ko‘rishi mumkin. Bizning kuzatishlarimizga ko‘ra, BiLSTM modeli ba’zan aynan shu sababli xato qilgan – chunki embeddinglar ko‘plab shakllarni to‘liq qamrab olmagan bo‘lishi mumkin, lug‘atda bo‘lmagan so‘zlar (OOV) muammosi chiqqan. mBERT

biroz buni yengadi, chunki subword tokenizatsiya bilan ilgari ko‘rmagan so‘zlarni ham bo‘lib tushunadi.

Yana bir qiyinchilik – neytral baholarni aniqlash. Neytral fikrlar odatda hissiyot bildiruvchi so‘zlarni o‘z ichiga olmaydi, asosan fakt yoki betaraf gap bo‘ladi. Masalan: *"Menyuda vegetarian taomlar ham bor ekan."* Bu gap neytral axborot. Modellar ko‘pincha bunday gaplarni neytral deb topdi. Ammo ba’zida *ton yoki kontekst* neytral gapga biroz hissiyot beradi: *"Menyuda vegetarian taomlar ham bor ekan, lekin tanlov kam."* Bunda birinchi qismini neytral deb, ikkinchi qism salbiy ohangda. Agar aspekt kesimlari noto‘g‘ri ajratilsa, model chalkashishi mumkin. Biz aspektlarni to‘g‘ri segmentlarga ajratganimiz uchun, har bir segment ancha aniq edi. Ammo impplitsit hissiyot (yoshiq ifodalangan norozilik yoki mamnunlik) modellarga qiyinchilik tug‘dirdi.

Ziddiyatli (qarama-qarshi) sinf eng qiyin bo‘ldi: hamma model bu sinfga past F1 ko‘rsatkich oldi (taxminan 0.40–0.50 orasida). Chunki juda kam misollar bo‘lgani sababli model bu holatni deyarli o‘rganmadi yoki noto‘g‘ri boshqa sinfga qo‘shib yubordi. Aslida ziddiyatli holatni aniqlash uchun modelga birgina gap ichida ijobiy va salbiy iboralar borligini payqash lozim. mBERT ba’zan buni to‘g‘ri qilganini ko‘rdik: masalan, *"Xizmat zo‘r edi, lekin ovqat yaxshi emas"* jumlasini mBERT aspektlar bo‘yicha qarama-qarshi deb belgilashga yaqinlashdi (ya’ni xizmatga ijobiy, ovqatga salbiy alohida), lekin biz hozir ushbu vazifani alohida aspekt kesimlariga ajratganimiz uchun, modelning vazifasi aslida bitta kesimda ikkita hissiyotni topish emas edi. Shu sabab, bizning formulamizda *ziddiyatli* sinf kam bo‘lgani uchun, uning past natijasi umumiy ishlashga katta ta’sir ko‘rsatmadi.

Xulosa qilish mumkinki, modellar ijobiy vs salbiyni yaxshi farqlay oladi, chunki bu bo‘yicha so‘zlar va ifodalar aniq. Neytral va ziddiyatlini farqlash esa ancha nozik til tushunishni talab qiladi. Kelajakda modellarni yaxshilash uchun O‘zbek tilining o‘ziga xos lug‘aviy va grammatik xususiyatlarini yanada hisobga olish zarur. Masalan, morfologik tahlil yordamida so‘zning negizini ajratish va qo‘shimchalarni alohida ishlov berish, yoki emotion lexicon (so‘zlarning hissiy lug‘ati) tuzib, modellarga qo‘shimcha bilim berish mumkin. Tadqiqotlarda tilning o‘ziga xos tomonlarini inobatga olish natijani sezilarli yaxshilashi mumkinligi qayd etilgan. O‘zbek tili uchun ham shunday yondashuvlar samara berishi kutiladi.

Xulosa va kelajakdagi ishlar

Ushbu ish doirasida biz O‘zbek tili uchun aspektga asoslangan hissiyot tahlili (ABSA) korpusini yaratdik va unda boshlang‘ich bir nechta modelni sinovdan o‘tkazdik. Korpus 7 mingdan ortiq belgilangan aspekt terminlari va 5 ta umumiy aspekt kategoriyasini o‘z ichiga olgan bo‘lib, har bir aspekt bo‘yicha ijobiy, salbiy, neytral yoki ziddiyatli baholar mavjud. Bu O‘zbek tilida birinchi keng qamrovli ABSA resursi sifatida xizmat qiladi, deb hisoblashimiz mumkin. Tajribalarda shuni

ko‘rdikki, zamonaviy model – ko‘p tilli BERT – O‘zbek tilidagi nozik hissiyot farqlarini aniqlashda juda yaxshi natija berdi va an’anaviy SVM hamda oddiy BiLSTM modellaridan ustun keldi. Bu shuni tasdiqlaydiki, transformer arxitekturalari hatto resurslari cheklangan tillarda ham katta foyda keltiradi, ayniqsa ular uchun ma’lumot to‘plamini boyitish imkoniyati mavjud bo‘lsa. Shu bilan birga, klassik yondashuvlar ham ma’lum darajada bazaviy natija ko‘rsatib, ayrim holatlarda qoniqarli ishlashi mumkinligini kuzatdik – bu esa amaliy loyihalarda resurs kam bo‘lsa ham tezkor yechim sifatida qo‘llash mumkinligini bildiradi.

Kelajakdagi ishlar bir necha yo‘nalishda olib borilishi mumkin. Birinchidan, korpusning o‘zini kengaytirish va boyitish muhim. Hozirgi korpus faqat restoran va ovqatlanish joylari sharhlarini qamrab oldi; kelajakda uni boshqa domenlar (masalan, texnika mahsulotlari sharhlari, kitob yoki film izohlari, ijtimoiy tarmoqdagi postlar) bilan kengaytirish rejalashtirilgan. Bu orqali model boshqa mavzulardagi aspektlarni ham o‘rgana oladi va umumlashgan ko‘nikma hosil qiladi. Korpus hajmining oshishi, ayniqsa neytral va ziddiyatli misollar sonini ko‘paytirish, modelning bu sinflarni yaxshiroq farqlashiga yordam beradi.

Ikkinchidan, hozirgi ishda biz aspektlar oldindan belgilangan holda ularning sentimentini klassifikatsiya qildik. Kelgusida ABSAning yanada qiyin shakli – birlashtirilgan aspekt-opinion ajratish va polaritetni aniqlash vazifasini ko‘rib chiqish zarur. Ya’ni, modelga faqat matn berilgan holda, u o‘zi aspekt terminlarini ajratib va ularga tegishli bahoni biryo‘la belgilashi kerak bo‘ladi. Bu End-to-End ABSA modeli bo‘lib, zamonaviy yondashuvlarda masalan seq2seq yoki pointer network usullar bilan hal qilinishi mumkin. Bunday model bizning korpus asosida o‘rgatilsa, real hayotdagi yangi sharhlarni avtomatik tahlil qilish imkoniyati paydo bo‘ladi (masalan, restoran haqidagi yangi izohdan avtomatik ravishda “ovqat – ijobiy, xizmat – salbiy” degan xulosa chiqarish).

Uchinchidan, ko‘p tillik va ko‘ppmodal yondashuvlar ham qiziqarli yo‘nalishdir. O‘zbek tili uchun resurslar oz bo‘lgani bois, cross-lingual transfer (tillararo o‘tkazma) usullaridan foydalanish samara beradi. Masalan, Turk, Ozarboyjon yoki Rus tilidagi o‘xshash ABSA korpuslaridan modelga qo‘shimcha trening uchun foydalanish yoki oldindan o‘rgatilgan XLM-RoBERTa kabi yanada kuchli ko‘p tilli modelni sinab ko‘rish mumkin. Ba’zi tadqiqotlarda yaqin tillarni birgalikda o‘qitish past resursli til natijalarini yaxshilashini ko‘rsatgan. Shuningdek, hozir paydo bo‘layotgan ChatGPT, GPT-4 kabi katta dastlabki modellar yordamida O‘zbek tilida aspekt-sentiment belgilangan sun’iy misollar yaratish va modelni few-shot tarzda o‘rgatish usullarini ham tekshirish mumkin – bu biz ilgari eslatgan kam resurs muammosini yengillashtirishga xizmat qiladi

Nihoyat, ABSA natijalarini real qo‘llashga tatbiq etish ham muhim yo‘nalish: masalan, O‘zbek tilidagi onlayn do‘konlar sharhlarini tahlil qilib, qaysi

jihatlar bo‘yicha mijozlar roziligi yoki noroziligi ko‘proq ekanini aniqlash, yoki mehmonxonalar uchun sharhlardagi muammoli jihatlarni (tozalik, qulaylik, narx) avtomatik ravishda monitor qilish. Bizning korpus va bazaviy modellar shu kabi amaliy loyihalar uchun boshlang‘ich poydevor vazifasini o‘tashi mumkin.

Xulosa qilib aytganda, O‘zbek tilida aspektga asoslangan hissiyot tahlili sohasida ilk qadamlar qo‘yildi va bu ishlanmalar hali ko‘p takomillashtirishlar va izlanishlarga ochiq. Ushbu tadqiqot natijalari shuni ko‘rsatadiki, ona tilimizdagi matnlardan yanada chuqurroq ma’lumot ajratib olish va foydali bilimlarga ega bo‘lish mumkin – buning uchun tilshunoslik bilimlari va zamonaviy sun’iy intellekt yondashuvlarini uyg‘un qo‘llash zarur bo‘ladi. Kelgusida korpusni kengaytirish, modellarni takomillashtirish va boshqa jamoalar bilan hamkorlikda ochiq platformalar yaratish orqali O‘zbek tilida ABSA va umuman NLP sohasini yangi bosqichga olib chiqish imkoni mavjud. Bu esa nafaqat ilmiy jamoaga, balki O‘zbek auditoriyasiga xizmat ko‘rsatuvchi ko‘plab tizimlar (masalan, onlayn sharh tahlilchilari, mijoz fikrini o‘rganuvchi dasturlar) rivojiga hissa qo‘shadi. Bizning ishimiz shu yo‘nalishda kichik bo‘lsa-da, muhim bir qadam bo‘lib, kelajakdagi tadqiqotlar va ishlanmalar uchun zamin yaratadi.

Foydalanilgan adabiyotlar:

1. Aziz, K., Ji, D., Chakrabarti, P. et al. Unifying aspect-based sentiment analysis BERT and multi-layered graph convolutional networks for comprehensive sentiment dissection. Sci Rep 14, 14646 (2024). https://doi.org/10.1038/s41598-024-61886-7 .

2. Jiahao Bu, Lei Ren and others, 2021. ASAP: A Chinese Review Dataset Towards Aspect Category Sentiment Analysis and Rating Prediction

3. Sanatbek Gayratovich Matlatipov, Jaloliddin Rajabov, Elmurod Kuriyozov, and Mersaid Aripov. 2024. UzABSA: Aspect-Based Sentiment Analysis for the Uzbek Language. In Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages @ LREC-COLING 2024, pages 394–403, Torino, Italia. ELRA and ICCL.

4. Zümberoğlu, K.B.; Dik, S.Z.; Karadeniz, B.S.; Sahmoud, S. Towards Better Sentiment Analysis in the Turkish Language: Dataset Improvements and Model Innovations. Appl. Sci. 2025, 15, 2062

5. Корпусы: аннотированные и неаннотированные. Лингвистическая аннотация (разметка) и метаданные. – Текст: электронный // Myfilology.ru – информационный филологический ресурс: [сайт]. – URL: https://myfilology.ru//177/korpusy-annotirovannye neannotirovannyelingvisticheskaya annotacziya-razmetka-i-metadannye/ (дата обращения: 2.12.2022).