IJTIMOIY TARMOQLARDAGI MATNLARNI SENTIMENT TAHLIL QILISH: BOSQICHMA-BOSQICH YONDASHUV

Elov Botir Boltayevich
ToshDO‘TAU Texnika fanlari bo‘yicha falsafa doktori PhD, dotsent
E-mail: elov@navoiy-uni.uz
Malika Suyunova Odil qizi
ToshDO‘TAU tayanch doktoranti
E-mail:malikasuyunova0@gmail.com

Annotatsiya. Ushbu maqolada o‘zbek tilidagi ijtimoiy tarmoqlardagi matnlarni aspektga asoslangan sentiment tahlil (ABSA) orqali tahlil qilish uchun bosqichma-bosqich yondashuv taklif etiladi. Tadqiqot ma’lumotlarni yig‘ish, tozalash, vektorlashtirish va modellashtirish jarayonlarini o‘z ichiga oladi. Shuningdek, ABSA jarayonida qo‘llaniladigan Naive Bayes, SVM, CRF, LSTM, BERT va RoBERTa kabi modellar va ularning afzalliklari, kamchiliklari ham tahlil qilingan. Maqolada o‘zbek tilidagi morfologik murakkabliklar, turli iboraviy ifodalar va aralash tillar kabi qiyinchiliklar ham ko‘rib chiqilgan. Tadqiqot natijalari turli vizualizatsiya usullari bilan taqdim etilib, ijtimoiy tarmoqlardagi sharhlar asosida jamiyat kayfiyatini aniqlash va foydalanuvchi fikrlarini baholash imkonini beradi.

Kalit so‘zlar: ABSA, ma’lumotlarni yig‘ish va tozalash, aspekt ajratish, sentiment aniqlash, Naive Bayes, SVM, CRF, LSTM, BERT, RoBERTa

Abstract. This article proposes a step-by-step approach to analyzing Uzbek-language social media texts using Aspect-Based Sentiment Analysis (ABSA). The study covers data collection, cleaning, vectorization, and modeling processes. It also examines the models commonly used in ABSA, such as Naive Bayes, SVM, CRF, LSTM, BERT, and RoBERTa, along with their advantages and limitations. The article discusses linguistic challenges specific to the Uzbek language, including morphological complexity, idiomatic expressions, and code-switching. The research findings are presented through various visualization techniques, enabling the assessment of public sentiment and user opinions based on social media reviews.

Keywords: ABSA, data collection and cleaning, aspect extraction, sentiment detection, Naive Bayes, SVM, CRF, LSTM, BERT, RoBERTa.

Аннотация. В данной статье предлагается пошаговый подход к анализу узбекоязычных текстов из социальных сетей с использованием аспектно-ориентированного анализа настроений (ABSA). Исследование охватывает этапы сбора данных, их очистки, векторизации и моделирования. Также рассматриваются модели, применяемые в процессе ABSA, такие как Naive Bayes, SVM, CRF, LSTM, BERT и RoBERTa, а также их преимущества и недостатки. В статье анализируются лингвистические сложности узбекского

языка, включая морфологическую сложность, фразеологические выражения и смешение языков. Результаты исследования представлены с помощью различных методов визуализации, что позволяет определять общественное настроение и оценивать мнения пользователей на основе комментариев из социальных сетей.

Ключевые слова: ABSA, сбор и очистка данных, выделение аспектов, определение тональности, Naive Bayes, SVM, CRF, LSTM, BERT, RoBERTa.

Kirish

So‘nggi yillarda ijtimoiy tarmoqlardagi foydalanuvchilarning fikri, his-tuyg‘usini, kayfiyatini o‘rganish asosiy mavzulardan biriga aylandi. Har kuni millionlab foydalanuvchilar turli mavzularda postlar, izohlar va fikrlar qoldiradi. Ushbu matnlar jamiyatdagi ijtimoiy kayfiyatni, mahsulot va xizmatlarga bo‘lgan munosabatni, shuningdek iqtisodiy va siyosiy jarayonlarga nisbatan ommaviy fikrni tahlil qilishda muhim manba hisoblanadi. Tabiiy tilni qayta ishlash (NLP) sohasida sentiment tahlil qilish muhim yo‘nalish sifatida qo‘llanilib kelmoqda. Sentiment tahlilning asosiy maqsadi – foydalanuvchi fikrlarini ijobiy, salbiy va neytral kabi avtomatik baholashdir. Dastlab bu yondashuv ingliz tilidagi matnlar asosida ishlab chiqilgan bo‘lsa-da, so‘nggi yillarda u turli tillarga, jumladan, **turkiy tillar**, xususan, **o‘zbek tili** uchun ham moslashtirilmoqda. Aspektga asoslangan sentiment tahlil (ABSA) matnni faqat ijobiy, salbiy va neytralga baholash bilan cheklanib qolmasdan, matndagi fikr bildirilayotgan aspekt va uning hissiy bahosini aniq qilib yoritib beradi. Shu sababli ABSA boshqa sentiment tahlil turlaridan aniqligi bilan ajralib turadi. Masalan, “Telefonning ishlash tezligi zo‘r ekan, lekin kamerasi tiniq emas” – ushbu gapni ABSA usulida sentiment tahlil qilinganda tezligi – zo‘r (ijobiy); kamera – tiniq emas (salbiy) kabi aniq natija chiqariladi. Shu tarzda ABSA yondashuvi fikrning aniq komponentlariga e’tibor qaratgan holda tahlilning aniqligini oshiradi.

O‘zbek tilidagi gaplarni ABSA usulida sentiment tahlil qilishda bir qancha qiyinchiliklar yuzaga kelishi mumkin. Chunki o‘zbek tili morfologik murakkabliklarga ega, so‘zlashuv uslubi xilma-xil, turli tillar aralashgan matnlar, ibora va kinoyalar turli murakkabliklarni keltirib chiqaradi. Shu sababli o‘zbek tili uchun bosqichma-bosqich yondashuv asosida sentiment tahlilni amalga oshirish muhim ahamiyat kasb etadi. Bunda ma’lumotlarni yig‘ish, tozalash, aspektni aniqlash va baholab yakuniy natijani chiqarish kabi bosqichlar har bir til xususiyatidan kelib chiqib moslashtirilishi lozim.

Mazkur maqolada ijtimoiy tarmoqlardagi o‘zbek tili matnlarini tahlil qilish uchun bosqichma-bosqich metodologik yondashuv taklif qilinadi. Ushbu yondashuv ma’lumotlarni to‘plashdan boshlab, model natijalarini baholashgacha bo‘lgan jarayonlarni o‘z ichiga oladi. Taklif etilgan yondashuv o‘zbek tilidagi

ISSN 3030-3907

ijtimoiy tarmoqlar (Telegram, Twitter, You Tube, Facebook va boshqalar) matnlarini avtomatik tahlil qilish orqali jamiyat kayfiyatini aniqlash, foydalanuvchilar fikrlarini baholash hamda savdo, ta'lim, siyosat, sog'liqni saqlash va boshqa bir qancha sohalarda kamchiliklarini aniqlash va ularni bartaraf etish uchun zamin yaratadi.

ADABIYOTLAR TAHLILI

Aspektga asoslangan sentiment tahlil bo'yicha jahon tilshunosligida bir qancha amalga oshirilgan ishlar mavjud. Jumladan, 1-jadvalda ABSA uchun ma'lumotlarni yig'ish va tozalash borasidagi amaliy ishlar qisqacha umumlashtirilib ko'rsatilgan:

1-jadval

ABSA uchun ma'lumotlarni yig'ish va tozalash borasidagi tadqiqotlar

M. Apidianaki va boshq. [1]	Ushbu tadqiqot fransuz tili uchun yaratilgan ABSA tizimlarini rivojlantirishda muhim o'rin tutadi. Tadqiqotchilar SemEval-2016 doirasida ko'p tilli (multilingual) platforma asosida turli domenlarda (restoran, muzey) ma'lumotlar to'plamlarini ishlab chiqqan. Mazkur ish ABSA tizimlarining ko'p domenli sinov muhiti uchun asos bo'lgan. Shu bilan birga, u turkiy tillar, ayniqsa o'zbek tili uchun ham shunday parallel korpuslarni yaratish zarurligini ko'rsatadi.
YR Regatte va boshq. [2]	Tadqiqot past resursli tillar (low-resource languages) uchun ABSA ma'lumotlar to'plamini yaratish metodologiyasini taklif etgan. Telugu tili uchun ishlab chiqilgan bu dataset, annotatsiya jarayoni va baholash mezonlari bilan ahamiyatlidir. Mazkur ish resurs yetishmasligi muammosiga yechim taklif etib, o'zbek tili uchun ham ABSA korpuslarini yaratilishida namuna bo'la oladi. Xususan, annotatsiya jarayonida inson omilining ahamiyatini yoritib, ko'p bosqichli validatsiya yondashuvini taqdim etadi.
Chebolu va boshq. [3]	Tadqiqotchilar mavjud ABSA datasetlarining keng tahlilini o'tkazib, 25 dan ortiq domenni o'z ichiga olgan 98 ta ochiq korpusni o'rganganlar. Ushbu ishda datasetlarning afzallik va cheklovlari batafsil tahlil qilingan bo'lib, kelgusida yangi korpuslar yaratishda sifat, balans va domen xilma-xilligi muhimligini ta'kidlaydi. Mazkur tahlil o'zbek tili uchun korpus yaratishda asosiy metodologik mezonlarni belgilashga yordam beradi.

ISSN 3030-3907

Chengyan Wu va boshq. [4]	Mualliflar tomonidan taqdim etilgan M-ABSA korpusi hozircha eng keng ko‘lamli ko‘p tilli dataset hisoblanadi (21 til, 7 domen). Har bir matn aspekt termini, kategoriya va polaritet (triplet) formatida belgilangan. Bu yondashuv ABSA uchun umumlashtirilgan multilingval model yaratish imkonini beradi. Mazkur ish o‘zbek tilini ham global ABSA tadqiqotlariga integratsiya qilish uchun yo‘l ochuvchi namuna sifatida muhimdir.
Hua va boshq. [5]	Tadqiqot ta’lim sohasi uchun maxsus yaratilgan EduRABSA korpusini taqdim etadi. Ushbu dataset kurslar, o‘qituvchilar va universitetlar haqidagi talaba fikrlarini qamrab olgan bo‘lib, aspektlarni sohaga xos ravishda aniqlash tamoyilini ilgari suradi. Bu yondashuv ABSA’ni domen-spetsifik tahlilga moslashtirish imkonini beradi. O‘zbek tilida, xususan onlayn ta’lim yoki MOOC platformalari uchun ham shu metodologiyani qo‘llash katta istiqbolga ega.

Aspektga asoslangan sentiment tahlil (ABSA) jarayonida ma’lumotlarni yig‘ish muhim bosqich bo‘lib, u tizimning ishlash samaradorligini bevosita belgilaydi. Tadqiqotlarda ma’lumotlarni yig‘ishning bir nechta usullari qo‘llanilgan, jumladan: **qo‘lda annotatsiya qilish**, **crowdsourcing**, hamda **yarim avtomatik texnikalar** (masalan, bootstrapping va active learning). Tanlangan yondashuvlar mavjud resurslar va tadqiqot maqsadlariga bog‘liq.

Jahon miqyosida ABSA uchun bir qator benchmark datasetlar yaratilgan. Masalan, Pontiki va boshq. (2014) “SemEval-2014 Task 4” tadqiqotida ingliz tilidagi restoran va noutbuk sharhlari asosida dastlabki benchmark dataset ishlab chiqilgan. Tadqiqotchilar POS-tag va qo‘lda annotatsiya yordamida aspektlarni va ularning polaritetini belgilashgan. Shu asosda, keyinchalik, SemEval-2015/2016 datasetlari yaratilgan va ularning yordamida feature-engineering, pipeline yondashuvlari, hamda bir qator klassik mashinaviy o‘rganish metodlari (SVM, CRF) sinovdan o‘tkazilgan [5, 6].

Bundan tashqari, global tadqiqotlarda turli tizimlar uchun to‘liq ABSA zanjiri (end-to-end) baholangan. Masalan, C. Brun va V. Nikoulina Foursquare sharhlari asosida yangi ABSA datasetini yaratishgan va tizimlarning samaradorligini tahlil qilishgan [7]. Sun va boshqalar esa ABSA vazifasini savol-javob (QA) va tabiiy til asosidagi xulosalash (NLI) muammolariga o‘xshatib, BERT modelini fine-tuning yordamida SentiHood va SemEval-2014 Task 4 datasetlarida ilg‘or natijalarga erishganlar [8]. Sampathirao va boshqalar ABSAning ahamiyati, muammolari va turli sohalardagi qo‘llanilish imkoniyatlarini keng qamrovli ravishda o‘rgangan. Shuningdek, ABSAni

ISSN 3030-3907

bosqichma-bosqich amalga oshirish tartibini hamda so‘nggi modellar va ma’lumotlar to‘plamlarini (datasets) tahlil qilib, ularning afzalliklari, cheklovlari va kelajakdagi yo‘nalishlarini yoritishgan [9].

Turkiy va o‘zbek tillarida ABSA uchun ma’lumotlarni yig‘ish bo‘yicha bir qator tadqiqotlar amalga oshirilgan. Masalan, Bayraktar va boshq. turk tilidagi restoran sharhlari asosida rule-based va statistik yondashuvlar yordamida ABSA datasetini yaratgan va lingvistik metodlarni qo‘llashgan. Shu tadqiqot turk tilidagi dastlabki ABSA ishlaridan biri hisoblanadi [10]. Shuningdek, Bilgin va Turan ham turk tilida ABSA vazifasini token va gap darajasida klassifikatsiya yondashuvi bilan bajarishgan. Bu yondashuv turk tilida aspektlarni aniqlash va ularning sentimentini belgilash bo‘yicha muhim tajriba sifatida xizmat qiladi [11].

O‘zbek tilida esa, Matlatipov va boshq. (UzABSA) restoran sharhlari asosida dastlabki ABSA datasetini ishlab chiqqanlar. Ushbu ish qo‘lda annotatsiya va domen-spetsifik aspektlarni belgilash orqali ABSA jarayonini o‘zbek tiliga moslashtirish imkonini beradi [12].

Yuqorida keltirilgan tadqiqotlar 2-jadvalda jamlanib, ularning ma’lumot manbalari, asosiy yondashuvlari va natijalari ko‘rsatilgan. Bu jadval ABSA uchun ma’lumot yig‘ish bosqichidagi turli tillar va domenlardagi yondashuvlarni solishtirishga yordam beradi.

2-jadval

ABSA uchun ma’lumot yig‘ish va annotatsiya yondashuvlari bo‘yicha tadqiqotlar

Muallif(lar)	Yil	Asar nomi	Ma’lumot manbasi	Asosiy yondashuv	Natija
Pontiki et al.	2014	SemEval-2014 Task 4: Aspect Based Sentiment Analysis	Restoran, Laptop	Rule-based, SVM, CRF	ABSA uchun benchmark yaratilgan
Brun & Nikoulina	2018	Aspect based sentiment analysis into the wild.	Foursquare sharhlari	End-to-end ABSA	Yangi real-domenli dataset
Sun et al.	2019	Utilizing BERT for aspect-based sentiment analysis via constructing	SentiHood, SemEval-2014	QA & NLI asosida fine-tuning	State-of-the-art natijalar

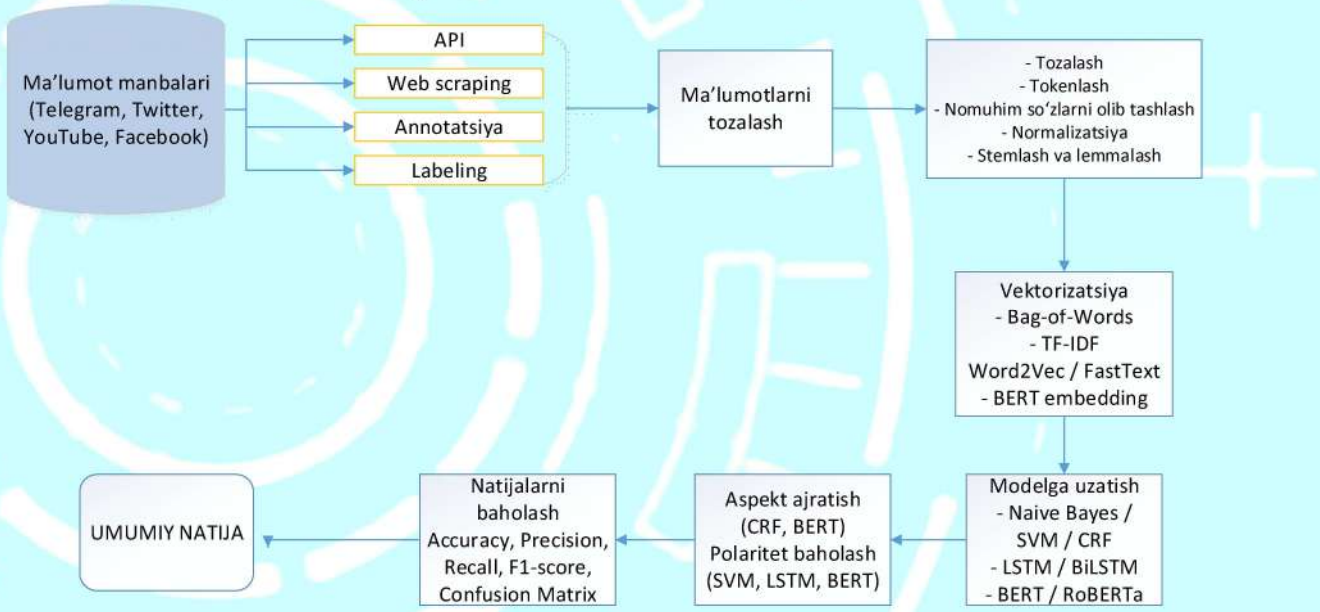
ISSN 3030-3907

		auxiliary sentence			
Sampathirao et al.	2022	Aspect-based sentiment analysis: A comprehensive survey of techniques and applications	Ko‘p domenli	Review (Survey)	Yondashuvlar va kelajak yo‘nalishlari
Metin Bilgin & Melek Turan	2021	Aspect-Based Sentiment Analysis for Turkish Reviews Using Token and Sequential Classification Methods	Turk tilidagi restoran sharhlari	Token-level (aspect term extraction) va sequence-level (sentiment classification) klassifikatsiya; transformer modellar (BERT, ConvBERT, ELECTRA, DistilBERT)	Aspect term va kategoriyalarni aniqlashda BERT eng yuqori aniqlik (92,20%); sentiment tahlilida ConvBERT eng yaxshi aniqlik (86,91%). Turk tilida nozik-granulyar ABSA uchun benchmark yaratildi.
Matlatipov S. va boshq. (UzABSA)	2024	Aspect-Based Sentiment Analysis for the Uzbek Language	O‘zbek tili restoran sharhlari	Qo‘lda annotatsiya, domen-spetsifik aspektlarni belgilash	O‘zbek tilida dastlabki ABSA datasetini yaratgan; ABSA jarayonini o‘zbek tiliga moslashtirgan

Aspektga asoslangan sentiment tahlilni amalga oshirish bosqichlari

ISSN 3030-3907

Ijtimoiy tarmoqlardagi sharhlar avval yig‘iladi va tozalanadi, qayta ishlangan matndagi aspektlar aniqlanadi va har bir aspekt uchun berilgan hissiy baho aniqlanadi, oxirida natijalar umumlashtirilib chiqariladi:



1-rasm. Aspektga asoslangan sentiment tahlili modellarining sxemasi

Ma’lumotlarni yig‘ish va tozalash

ABSA jarayonining eng muhim qismlaridan biri – ma’lumotlarni yig‘ishdir. Turli ijtimoiy tarmoqlardagi postlar, sharhlar to‘planadi va API yoki Web scraping orqali yig‘iladi. Yig‘ilgan matnlar tozalash bosqichiga o‘tadi. Tozalash bosqichida:

– Emoji va maxsus belgilarni olib tashlash – faqat lingvistik ma’lumot qoldiriladi;

– URL, @username va hashtaglarni tozalash;

– Tokenizatsiya – matnni so‘z yoki ibora bo‘laklariga ajratish;

– Stop-word removal – “va”, “ham”, “bu” kabi hissiy qiymati yo‘q so‘zlarni olib tashlash;

– Lemmatizatsiya yoki stemming – so‘zlarni asl (bazaviy) shaklga keltirish;

– Til aniqlash va aralash tillarni filtrlash – o‘zbek tiliga tegishli bo‘lmagan so‘zlarni chiqarib tashlash;

– Emojilarni semantik qiymat bilan almashtirish (masalan, 😊 → “ijobiy”).

Ma’lumotlar tozalanib vektorlashtirilgach Naive Bayes, Support Vector Machine (SVM), Conditional Random Fields (CRF), Long Short-Term Memory (LSTM), BERT va RoBERTa kabi modellar yordamida sentiment tahlil qilinadi.

Naive Bayes modeli – bu Bayes teoremasiga asoslangan matnni avtomatik tarzda ma’lum sinflarga ajratuvchi ehtimollik klassifikatoridir. Bunda matn tozalanib Bag-of-Words (BoW) yoki TF-IDF (Term Frequency-Inverse Document Frequency) yondashuvlaridan foydalanib raqamli shaklga o‘tkaziladi. Ushbu bosqichda har bir so‘zning chastotasi yoki muhimligi hisoblanib, matn vektor ko‘rinishida ifodalanadi. Masalan, *“Bu film ajoyib edi”* jumlasi uchun model so‘zlarning chastotasini aniqlaydi: *film=1, ajoyib=1, yomon=0* va hokazo.

ISSN 3030-3907

Modelning aniqligini baholash uchun turli metrikalar – **accuracy**, **precision**, **recall** va **F1-score** hisoblanadi. Ushbu ko‘rsatkichlar modelning real ma’lumotlar ustida qanchalik samarali ishlashini tahlil qilishga yordam beradi. Baholash jarayonida kross-valikatsiya usuli qo‘llanilib, modelning umumlashma qobiliyati tekshiriladi. Naive Bayesning asosiy afzalliklari – kichik hajmdagi ma’lumotlar bilan ham yuqori aniqlikka erishish, tez o‘qitilish va matn tahlilida boshlang‘ich klassifikator sifatida qulaylik yaratishdir.

SVM – bu chiziqli va no-chiziqli ajratish gipertekisliklari orqali sinflar o‘rtasidagi maksimal chegarani topishga asoslangan nazoratli o‘rganish (supervised learning) algoritmidir. Modelning asosiy g‘oyasi – ma’lumotlarni vektor fazosida har bir sinf uchun chegaraviy nuqtalar (support vectors) orasidagi masofa maksimal bo‘ladi. ABSA kontekstida SVM aspekt polaritetini aniqlash uchun keng qo‘llaniladi. Masalan, foydalanuvchi fikrida *“xizmat sifati”* va *“narx”* aspektlariga oid so‘zlar vektorlarga aylantirilib, TF-IDF yoki Word2Vec yordamida xususiyat sifatida kiritiladi. Model bu vektorlarni chiziqli kernel (Linear), radial asosli kernel (RBF) yoki polinomial kernel yordamida sinflarga ajratadi. Afzalligi shundaki, yuqori aniqlikka ega, katta o‘lchamli vektor fazolarida samarali ishlaydi, overfittingga nisbatan chidamli. Kamchiligi esa parametrlarni moslashtirish murakkab va ma’lumotlar katta bo‘lganda hisoblash xarajati yuqori bo‘ladi.

CRF – berilgan ketma-ketlikdagi belgilarni birgalikda tahlil qilib, ularning bir-biriga bog‘liqligini hisobga olgan holda eng mos belgilar ketma-ketligini tanlashdir. Masalan, *“Telefonning kamerasi ajoyib”* degan fikrda CRF *“kamerasi”* so‘zini **aspekt termini** sifatida aniqlay oladi, chunki u “Telefon” so‘zi bilan sintaktik jihatdan bog‘langan va kontekstual ma’noda ijobiy baholangan. Afzalligi so‘zlar o‘rtasidagi bog‘liqlikni hisobga oladi, ketma-ketlikdagi kontekstual ma’noni to‘liq tahlil qiladi, Aspekt ajratish bosqichida yuqori aniqlik beradi. Kamchiligi esa qo‘lda belgilanadigan xususiyatlar ko‘p mehmat qiladi va o‘qitish jarayoni sekin.

LSTM – bu recurrent neural network (RNN) ning takomillashtirilgan turi bo‘lib, u vaqt ketma-ketliklaridagi uzoq muddatli bog‘lanishlarni (long-term dependencies) saqlab qolishga qodir. LSTM tarmog‘i “input gate”, “forget gate” va “output gate” mexanizmlari orqali har bir so‘zning kontekstini dinamik ravishda yangilab boradi. Bu model, ayniqsa, matndagi murakkab sintaktik va semantik bog‘liqliklarni o‘rganishda samarali natija beradi. Masalan, *“Oshxonaning taomi yaxshi, lekin xizmat sekin”* jumlasida LSTM *“taom”* uchun **ijobiy**, *“xizmat”* uchun **salbiy** baho berishni o‘rganadi. Afzalligi uzoq kontekstni hisobga oladi, annotatsiyasiz katta matnlarda ham o‘rganishga mos, semantik tahlilda yuqori

ISSN 3030-3907

samaradorlikka ega. Kamchiligi esa hisoblash murakkabligi yuqori va o‘qitish uchun katta hajmdagi ma’lumot talab etiladi.

BERT (Bidirectional Encoder Representations from Transformers) – transformer arxitekturasiga asoslangan chuqur o‘rganish modeli bo‘lib, u matnni ikki yo‘nalishda (chapdan o‘ngga va o‘ngdan chapga) o‘qib, kontekstual ma’noni chuqur tushunishga imkon beradi.

RoBERTa (Robustly Optimized BERT Approach) esa BERTning optimallashtirilgan versiyasi bo‘lib, ko‘proq ma’lumot bilan, dinamik masklash va o‘zgartirilgan o‘qitish strategiyasi yordamida yuqori natijalarga erishadi. ABSA sohasida BERT va RoBERTa modellari aspekt aniqlash, sentiment klassifikatsiyasi va aspekt-segment kontekst bog‘lanishini tahlil qilish bosqichlarida eng samarali yondashuvlar sifatida tan olingan. Masalan, “Narx yaxshi, lekin xizmat yomon” gapida model “narx” va “xizmat” aspektlarini ajratib, ularning har biriga nisbatan sentimentni kontekst asosida to‘g‘ri aniqlaydi. Afzalliklari kontekstni ikki yo‘nalishda chuqur o‘rganadi, aspekt va sentiment o‘rtasidagi murakkab bog‘lanishlarni tushunadi, oldindan o‘qitilgan (pre-trained) modellar asosida tez moslashadi. Kamchiligi esa juda katta hisoblash resurslari talab qiladi va o‘qitish vaqti uzun.

Aspektlarni aniqlash

Aspekt tahlili – matnda fikr bildirilayotgan obyekt yoki mavzuni aniqlash jarayonidir. Ushbu bosqichda quyidagi yondashuvlar qo‘llanadi:

- Rule-based usullar: lug‘at va qoidalar asosida aspektlarni aniqlash;
- Statistik usullar: TF-IDF, PMI yoki frequency-based tahlil orqali asosiy aspektlarni topish;
- Mashinalar orqali o‘rganish: CRF, LSTM va transformer modellar (masalan, BERT) yordamida avtomatik aspekt ajratish;

Masalan, *“Telefonning ishlash tezligi zo‘r, lekin kamerasi tiniq emas”* gapida *tezligi* va *kamera* so‘zlari **aspekt** sifatida aniqlanadi, bu esa keyingi sentiment bahosini aniqlashda ishlatiladi.

Sentimentni aniqlash

Aniqlangan har bir aspekt uchun hissiy bahoni belgilash – ijobiy, salbiy yoki neytral – ushbu bosqichning asosiy maqsadidir. Bu yerda quyidagi yondashuvlar qo‘llanadi:

- Klassik mashinaviy o‘rganish usullari: SVM, Naive Bayes, Random Forest.
- Transformer modellar: BERT, RoBERTa, XLM-R va ularning fine-tuning variantlari orqali o‘zbek tiliga moslashtirilgan sentiment aniqlash.
- Polarity triplet yaratish: har bir aspekt uchun (aspect term, category, polarity) formatida natija saqlanadi.

Natijalarni baholash va integratsiya qilish

ISSN 3030-3907

Bu yakuniy bosqich hisoblanib, ABSA tizimi tomonidan olingan chiqish natijalarini tahlil qilish, ularning aniqligini mezonlar orqali baholash hamda turli modellar yoki manbalardan olingan ma’lumotlarni yagona tizimga birlashtirish jarayonidir. Bunda quyidagi ishlar amalga oshiriladi:

- Baholash mezonlari: Accuracy, Precision, Recall va F1-score kabi standart ko‘rsatkichlar yordamida model natijasi tekshiriladi.
- Vizualizatsiya: Word cloud, heatmap va sentiment trend grafikalari orqali aspekt va sentiment natijalari vizual tarzda taqdim etiladi.
- Integratsiya: Aspektlar va ularning sentimenti yagona tizimda jamlanib, ijtimoiy tarmoqlardagi umumiy kayfiyatni tahlil qilish imkoniyati yaratiladi.

XULOSA

Tadqiqot o‘zbek tilidagi ijtimoiy tarmoqlardagi matnlarni ABSA orqali tahlil qilish uchun bosqichma-bosqich yondashuvni qo‘llaydi. Ma’lumotlar Telegram, Twitter, YouTube va Facebook kabi platformalardan turli domenlarga (restoran, elektronika, ta’lim) oid postlar va sharhlardan yig‘iladi. Yig‘ilgan ma’lumotlar preprocessing jarayonidan o‘tadi: emoji va maxsus belgilarni semantik qiymat bilan almashtirish, URL va hashtaglarni tozalash, tokenizatsiya, stop-word removal, lemmatizatsiya, til aniqlash va vektorlashtirish. Tahlil jarayonida Naive Bayes, SVM, CRF, LSTM, BERT va RoBERTa kabi klassik va chuqur o‘rganish modellari qo‘llaniladi. Modellarning samaradorligi Accuracy, Precision, Recall va F1-score kabi mezonlar orqali baholanadi. Olingan natijalar vizualizatsiya qilinib (Word cloud, heatmap, sentiment trend) umumiy tizimga integratsiya qilinadi, bu orqali ijtimoiy tarmoqlardagi jamiyat kayfiyati va soha-spetsifik fikrlar aniqlanadi. Tadqiqot davomida foydalanuvchilarning shaxsiy ma’lumotlari himoyalanadi va annotatsiya jarayonida validatsiya bosqichlari qo‘llanadi. Taklif qilingan yondashuvlar o‘zbek tilidagi sentiment tahlilini rivojlantirish va turli sohalarda foydalanuvchi fikrlarini samarali tahlil qilish uchun asos bo‘lib xizmat qiladi.

Sentiment tahlil borasida o‘zbek tilishunosligida katta tadqiqot ishlari amalga oshirilmoqda. Jumladan, B.Elov va A. Abdullayev sentiment tahlil uchun katta hajmdagi datasetni yaratish jarayonini uch bosqichli tizim asosida tahlil qilishgan. Ushbu tadqiqot sentiment tahlil uchun katta hajmdagi datasetni yaratishning ilmiy asoslangan bosqichlarini taqdim etib, sohada innovatsion yechimlarni ilgari surgan[13].

FOYDALANILGAN ADABIYOTLAR RO‘YXATI

1. Apidianaki, M., Tannier, X., & Richart, C. (2016, January). Datasets for aspect-based sentiment analysis in French. In International Conference on Language Resources and Evaluation.

2. Regatte, Y. R., Gangula, R. R. R., & Mamidi, R. (2020, May). Dataset creation and evaluation of aspect based sentiment analysis in Telugu, a low resource

ISSN 3030-3907

language. In Proceedings of the Twelfth Language Resources and Evaluation Conference (pp. 5017-5024).
3. Chebolu, S. U. S., Dernoncourt, F., Lipka, N., & Solorio, T. (2023, November). A review of datasets for aspect-based sentiment analysis. In Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 611-628).
4. Wu, C., Ma, B., Liu, Y., Zhang, Z., Deng, N., Li, Y., ... & Plank, B. (2025). M-ABSA: A Multilingual Dataset for Aspect-Based Sentiment Analysis. arXiv preprint arXiv:2502.11824.
5. Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 27–35, Dublin, Ireland. Association for Computational Linguistics.
6. Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, et al.. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. 10th International Workshop on Semantic Evaluation (SemEval 2016), Jun 2016, San Diego, United States. pp.19 - 30, ⟨10.18653/v1/S16-1002⟩. ⟨hal-02407165⟩
7. Brun, C., & Nikoulina, V. (2018, October). Aspect based sentiment analysis into the wild. In Proceedings of the 9th workshop on computational approaches to subjectivity, sentiment and social media analysis (pp. 116-122).
8. Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. arXiv preprint arXiv:1903.09588.
9. Sampathirao Suneetha, S., & Row, V. (2023). Aspect-based sentiment analysis: A comprehensive survey of techniques and applications. J Data Acquis Process, 38(3), 177-203.10. Bayraktar, K., Yavanoglu, U., & Ozbilen, A. (2019, December). A rule-based holistic approach for Turkish aspect-based sentiment analysis. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 2154-2158). IEEE.
10. Metin Bilgin, Melek Turan. (2025). Aspect-Based Sentiment Analysis for Turkish Reviews Using Token and Sequential Classification Methods. (2025). International Journal of Computer and Information Technology(2279-0764), 14(3). https://doi.org/10.24203/2drscx54
11. Matlatipov, S. G., Rajabov, J., Kuriyozov, E., & Aripov, M. (2024, May). UzABSA: Aspect-Based Sentiment Analysis for the Uzbek Language.

ISSN 3030-3907

In Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages@ LREC-COLING 2024 (pp. 394-403).
12. Elov, B., & Abdullayev, A. (2025). SENTIMENT TAHLIL UCHUN KATTA HAJMDAGI DATASETNI YARATISH BOSQICHLARI. COMPUTER LINGUISTICS: PROBLEMS, SOLUTIONS, PROSPECTS, 1(1).

ISSN 3030-3907