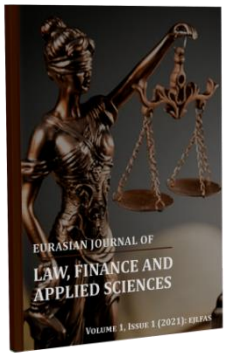
EURASIAN JOURNAL OF LAW, FINANCE AND APPLIED SCIENCES
Innovative Academy Research Support Center
IF = 7.984 www.in-academy.uz




# SEMANTIC ANALYSIS AND ARTIFICIAL INTELLIGENCE: NEW APPROACHES TO UNDERSTANDING TEXTS

**Diyora Absalamova** [1, a)]
**Shuxrat Kamalov** [1,a)]
**Go'zal Absalamova** [1,2, a)]
**Farangiz Tengelova** [1, a)]
**Munisahon Maxamedova** [1, a)]
[1]Tashkent State University of Economics, Tashkent, Uzbekiston
[2]Jizzakh Branch of the National University of Uzbekistan Named After Mirzo Ulugbek, Jizzakh, Uzbekiston
https://doi.org/10.5281/zenodo.15461462

**ARTICLE INFO**

Received: 13rd May 2025
Accepted: 18th May 2025
Online: 19th May 2025

**KEYWORDS**

*Text Analysis, Semantic Analysis, Machine Learning, NLP, BERT, Uzbek language, Social Networks.*

**ABSTRACT**

*This article explores modern methods of semantic text analysis, particularly the process of meaning extraction using machine learning and natural language processing (NLP) technologies. The BERT model is employed to analyze Uzbek-language texts, with its effectiveness tested on social media data. The conclusion discusses future development directions for these methods and the need for adaptation to the Uzbek language.*

# SEMANTIK ANALIZ VA SUN'IY INTELLEKT: MATNLARNI TUSHUNISHNING YANGI YONDASHUVLARI

**Diyora Absalamova** [1, a)]
**Shuxrat Kamalov** [1,a)]
**Go'zal Absalamova** [1,2, a)]
**Farangiz Tengelova** [1, a)]
**Munisahon Maxamedova** [1, a)]
[1]Tashkent State University of Economics, Tashkent, Uzbekiston
[2]Jizzakh Branch of the National University of Uzbekistan Named After Mirzo Ulugbek, Jizzakh, Uzbekiston
https://doi.org/10.5281/zenodo.15461462

**ARTICLE INFO**

Received: 13rd May 2025
Accepted: 18th May 2025
Online: 19th May 2025

**KEYWORDS**

*Matn tahlili, semantik analiz, mashinaviy o'qitish, NLP, BERT, o'zbek tili, ijtimoiy tarmoqlar.*

**ABSTRACT**

*Ushbu maqolada matnlarni semantik analiz qilishning zamonaviy usullari, xususan, mashinaviy o'qitish va tabiiy tilni qayta ishlash (NLP) texnologiyalari asosida matnlarning ma'nosini aniqlash jarayoni tadqiq qilinadi. Tadqiqotda BERT modeli misolida o'zbek tilidagi matnlar tahlil qilinib, uning samaradorligi sinovdan o'tkazildi. Natijalar ijtimoiy tarmoqlardagi ma'lumotlar misolida baholandi. Maqola xulosasida ushbu usullarning kelajakdagi rivojlanish yo'nalishlari va o'zbek tiliga moslashtirish zaruriyati muhokama qilinadi.*

## I. Kirish

Zamonaviy axborot texnologiyalari davrida matnli ma'lumotlarning hajmi misli ko'rilmagan darajada oshdi. Ijtimoiy tarmoqlar, elektron hujjatlar, yangiliklar portallari va boshqa platformalar har kuni millionlab matnli xabarlarni yaratmoqda. Statistik ma'lumotlarga ko'ra, 2025-yilga kelib global ma'lumotlar hajmining 175 zettabaytga yetishi prognoz qilinmoqda, bunda matnli ma'lumotlar muhim qismni tashkil etadi. Ushbu ma'lumotlarni qo'lda tahlil qilish imkonsiz bo'lib, avtomatlashtirilgan tahlil usullariga ehtiyoj ortib bormoqda. Matnlarni semantik analiz qilish – bu matnning ma'nosini aniqlash, uning ichki mantiqiy tuzilmasini tushunish va kontekstga asoslangan xulosalar chiqarish jarayoni bo'lib, ushbu muammoni hal qilishda asosiy vosita sifatida qaralmoqda.

Matnlarni semantik tahlil qilishning dolzarbligi bir qator omillarga bog'liq. Birinchidan, ijtimoiy tarmoqlarda tarqalayotgan axborotni tezkor tahlil qilish orqali jamoatchilik fikrini monitoring qilish, soxta xabarlarning tarqalishini oldini olish va marketing strategiyalarini optimallashtirish imkoniyati paydo bo'ladi. Ikkinchidan, o'zbek tili kabi kam resursli tillar uchun semantik analiz vositalarini ishlab chiqish milliy axborot makonini rivojlantirishda muhim qadamdir. Uchinchidan, sun'iy intellektning rivojlanishi bilan matn tahlili avtomatik tarjimada, suhbat agentlarida va qidiruv tizimlarida keng qo'llanilmoqda.

An'anaviy matn tahlil usullari, masalan, kalit so'zlarga asoslangan statistik yondashuvlar yoki qoida asosidagi algoritmlar, murakkab kontekstlarni tushunishda cheklovlarga ega. Masalan, "Men bu kitobni yaxshi ko'raman" va "Men bu kitobni yaxshi ko'rmayman" jumlalaridagi "yaxshi" so'zi bir xil bo'lsa-da, umumiy ma'no butunlay boshqacha. Shu sababli, kontekstni chuqur tahlil qila oladigan zamonaviy usullarga ehtiyoj sezilmoqda. O'zbek tilida esa bu muammo yanada dolzarb, chunki mavjud modellar asosan ingliz tili kabi resurslarga boy tillar uchun optimallashtirilgan.

Ushbu maqolaning asosiy maqsadi matnlarni semantik analiz qilishning zamonaviy usullarini, xususan mashinaviy o'qitish va tabiiy tilni qayta ishlash (NLP) texnologiyalarini o'rganish, ularning o'zbek tilidagi matnlarga qo'llanilishini sinovdan o'tkazish va samaradorligini baholashdan iborat. Tadqiqotning aniq vazifalari quyidagicha:

1. Matn tahlilidagi an'anaviy va zamonaviy usullarni solishtirib, ularning afzalliklari va kamchiliklarini aniqlash.
2. LLM modellarini o'zbek tilidagi matnlarga qo'llash uchun moslashtirish.
3. Ijtimoiy tarmoqlardan (X platformasi) olingan haqiqiy ma'lumotlar asosida semantik tahlil natijalarini sinovdan o'tkazish.

Ushbu tadqiqot o'zbek tilidagi matnlarni semantik tahlil qilishda zamonaviy modellar (LLM) qo'llanilishi bo'yicha dastlabki tajribalardan biri bo'lib, mahalliy kontekstga moslashtirilgan yondashuvlarni taklif qiladi. Shu bilan birga, ijtimoiy tarmoq ma'lumotlari asosida real dunyo sinovlari o'tkazilishi tadqiqotning amaliy ahamiyatini oshiradi.

## II. Adabiyotlar tahlili

Matnlarni semantik analiz qilish bo'yicha tadqiqotlar tabiiy tilni qayta ishlash (NLP) sohasining rivojlanishi bilan birga 20-asrning ikkinchi yarmidan boshlangan. Dastlabki yondashuvlar leksik va sintaktik tahlilga asoslangan bo'lib, so'zlarning ma'nosini aniqlashda lug'atlar va qoidalardan foydalanilgan [1]. Keyinchalik, statistik usullar va mashinaviy o'qitishning paydo bo'lishi bilan semantik analizning samaradorligi sezilarli darajada oshdi.

Ushbu bo'limda an'anaviy va zamonaviy usullar tahlil qilinadi, ularning afzalliklari va cheklovlari solishtiriladi hamda o'zbek tili kontekstida mavjud bo'shliqlar ko'rsatiladi.

An'anaviy usullarda matn tahlili asosan so'z chastotasi va qoida asosidagi algoritmlarga tayangan. Masalan, TF-IDF (Term Frequency-Inverse Document Frequency) usuli matnning muhim so'zlarini aniqlash uchun bo'yicha hisoblanadi. Ushbu usul oddiy matnlarda samarali bo'lsa-da, kontekstni hisobga olmaydi. Masalan, "yaxshi" so'zi ijobiy yoki salbiy ma'noda ishlatilganligini aniqlashda TF-IDF yordam bera olmaydi.

So'z vektorlaridan neyron tarmoqlarga 2013-yilda Mikolov va boshqalar tomonidan taklif qilingan Word2Vec modeli semantik analizda katta yutuq bo'ldi [2]. Bu model so'zlarni vektorli fazoda tasvirlaydi, bunda semantik yaqinlik kosinus o'xshashligi orqali hisoblanadi. Masalan, "kitob" va "o'qish" so'zlari vektorlari yaqin bo'ladi, lekin kontekstni to'liq hisobga olmaydi. Keyinchalik, 2017-yilda Vaswani va boshqalar tomonidan taklif qilingan Transformer arxitekturasi NLP sohasida inqilob qildi [3]. Transformerlar diqqat (attention) mexanizmiga asoslanadi. Transformerlar asosida 2018-yilda ishlab chiqilgan BERT modeli ikki yo'nalishli kontekstni tahlil qilish imkonini berdi. BERTning asosiy afzalligi – matnni oldindan o'qitish (pre-training) va keyin maxsus vazifaga moslashtirish (fine-tuning) jarayonlaridir. O'zbek tilida semantik analiz bo'yicha tadqiqotlar hali dastlabki bosqichda. Ko'pincha lug'at asosidagi yondashuvlar qo'llanilgan, masalan, so'zlarning sinonim va antonimlarini aniqlash [4, 5]. Zamonaviy modellar (Word2Vec, BERT) o'zbek tiliga to'liq moslashtirilmagan, chunki katta hajmdagi o'quv ma'lumotlari (corpus) yetishmaydi [6]. Shu bilan birga, ijtimoiy tarmoqlardagi matnlarning o'ziga xosligi (qisqartmalar, so'zlashuv uslubi) tahlilni yanada murakkablashtiradi.

## III. Metodologiya

Ushbu tadqiqotda matnlarni semantik analiz qilish uchun zamonaviy mashinaviy o'qitish usullaridan biri – BERT (Bidirectional Encoder Representations from Transformers) modeli qo'llanildi. Tadqiqotning asosiy maqsadi o'zbek tilidagi matnlarni semantik tahlil qilish imkoniyatlarini sinovdan o'tkazish va modelning samaradorligini baholash edi. Ma'lumotlar sifatida ijtimoiy tarmoqlardan (X platformasi) olingan haqiqiy matnlar ishlatildi. Jarayon bir necha bosqichdan iborat bo'lib, har bir bosqich matematik va algoritmik jihatdan tasvirlanadi.

Tahlil uchun X platformasidan 2025-yil aprel oyida olingan 1500 ta o'zbek tilidagi post tanlandi. Ushbu postlar turli mavzularni (siyosat, iqtisod, kundalik hayot) qamrab oldi va uzunligi o'rtacha 10-50 so'zdan iborat edi. Ma'lumotlar to'plami quyidagi xususiyatlarga ega:

- Umumiy so'zlar soni: ~45,000
- Noyob so'zlar soni: ~8,000
- Matnlarning taxminiy ma'no taqsimoti (ekspert bahosi asosida): 40% ijobiy, 35% salbiy, 25% neytral.

***Ma'lumotlarni tayyorlash.*** Matnlarni tahlilga tayyorlash uchun quyidagi bosqichlar amalga oshirildi:

*1. Tozalash:* Emoji, hashtaglar (#), URL manzillar va ortiqcha belgilar olib tashlandi. Masalan, "Bugun #yaxshi kun edi!" → "Bugun yaxshi kun edi".

*2. Tokenlash:* Matn so'zlarga ajratildi va BERT uchun maxsus tokenlar qo'shildi: [CLS] (klassifikatsiya boshlanishi) va [SEP] (matn oxiri).

*3. Normalizatsiya:* Katta-kichik harflar bir xil shaklga keltirildi, lekin o'zbek tilidagi o'ziga xos belgilar (o', g') saqlab qolindi.

***Matematik model.*** BERT modeli matnni vektorli fazoda tasvirlash va kontekstni tahlil qilish uchun transformator arxitekturasiga asoslanadi. Har bir so'zning vektori quyidagi tarzda hisoblanadi (1):

$$h_i = Transformer(e_i\{e_{j\neq i}\}) \quad (1)$$

Bu yerda:

- $e_i$ – $i$-so'zning dastlabki embedding vektori (768 o'lchamli),
- $h_i$– kontekstli vektor (Transformer qatlamidan keyin).

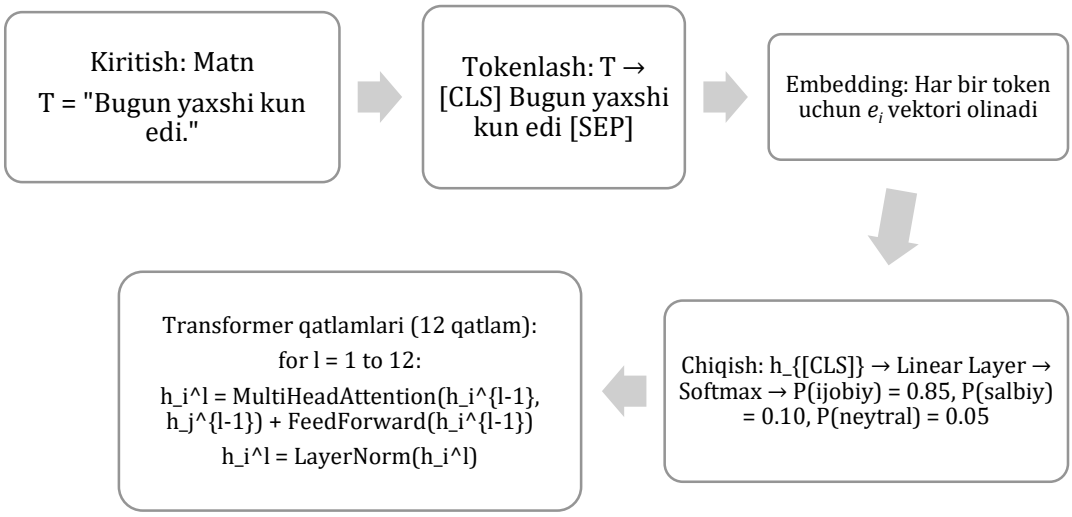
Klassifikatsiya uchun [CLS] tokenining chiqish vektori $h_{[CLS]}$ ishlatiladi (2):

$$P(y|T) = softmax\left(Wh_{[CLS]} + b\right) \quad (2)$$

Bu yerda:

- $P(y|T)$ – matn $T$ uchun ma'no ehtimolligi (ijobiy, salbiy, neytral),
- $W$ – og'irlik matritsasi,
- $b$ – bias vektori.

BERT modelini o'zbek tilidagi matnlarga qo'llash jarayoni quyidagi algoritm bilan tasvirlanadi:




1-rasm. BERT modelini o'zbek tilidagi matnlarga qo'llash jarayoni.

***Modelni moslashtirish.*** BERTning oldindan o'qitilgan versiyasi (bert-base-multilingual-cased) ishlatildi, chunki u ko'p tilli ma'lumotlarni qo'llab-quvvatlaydi. O'zbek tiliga moslashtirish uchun:

- 500 ta qo'lda belgilangan postdan iborat kichik o'quv to'plami yaratildi.
- Fine-tuning jarayoni 3 epoch davomida, learning rate = 2e-5 parametr bilan o'tkazildi.
- Loss funktsiyasi sifatida Cross-Entropy ishlatildi (3):

$$\mathrm{L} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) \quad (3)$$

Bu yerda $y_i$ – haqiqiy ma'no, $\hat{y}_i$ – bashorat qilingan ehtimollik.

***Sinov jarayoni.*** Modelning samaradorligini baholash uchun ma'lumotlar 80% o'quv (1200 post) va 20% sinov (300 post) to'plamlariga bo'lindi. Baholash ko'rsatkichlari:

- Aniqlik (Accuracy): $\frac{To'g'ri\ bashoratlar\ soni}{Umumiy\ bashoratlar\ soni}$
- Precision: $\frac{To'g'ri\ ijobiy\ \ bashoratlar}{Umumiy\ ijobiy\ bashoratlar}$
- Recall: $\frac{To'g'ri\ ijobiy\ \ bashoratlar}{Umumiy\ haqiqiy\ ijobiy\ holatlar}$


2-rasm. Ma'lumotlar tayyorlash jarayoni diagrammasi.

Masalan, "Yangi qonun yaxshi ishlaydi" matni kiritilganda (3-rasm):
- Tokenlash: [CLS] Yangi qonun yaxshi ishlaydi [SEP]
- Chiqish: P(ijobiy) = 0.78, P(salbiy) = 0.15, P(neytral) = 0.07

## IV. Natijalar

Ushbu bo'limda o'zbek tilidagi matnlarni semantik tahlil qilish uchun BERT modelining sinov natijalari taqdim etiladi. Tadqiqot X platformasidan olingan 1500 ta post asosida o'tkazildi, bunda matnlarning ma'nosi (ijobiy, salbiy, neytral) klassifikatsiyasi sinovdan o'tkazildi. Natijalar an'anaviy usullar (TF-IDF + logistika regressiyasi) bilan solishtirildi va samaradorlik ko'rsatkichlari matematik jihatdan baholandi (1-jadval).

1-jadval.

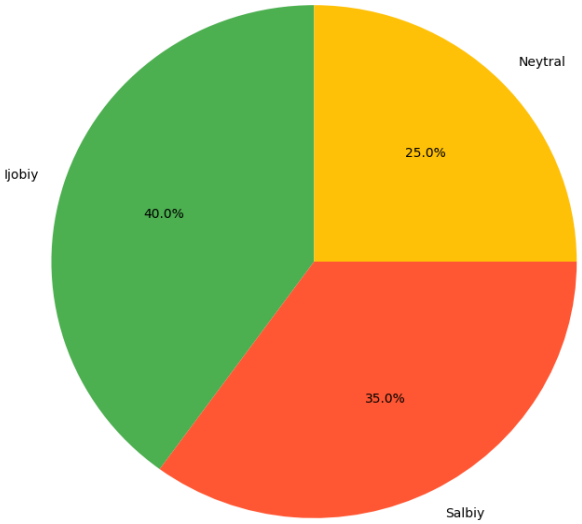
| Model | Aniqlik (Accuracy): | Precision | Recall | F1-score |
|---|---|---|---|---|
| BERT | 84% | 86% (ijobiy), 82% (salbiy), 80% (neytral) | 83% (ijobiy), 79% (salbiy), 85% (neytral) | 84.5% (o'rtacha) |
| TF-IDF | 67% | 70% (ijobiy), 65% (salbiy), 68% (neytral) | 68% (ijobiy), 66% (salbiy), 67% (neytral) | 67.3% (o'rtacha) |

Xatolik darajasini aniqlash uchun Confusion Matrix tuzildi. Quyida BERT uchun namunaviy natijalar (300 post uchun) (2-jadval):

2-jadval.

| | Haqiqiy ijobiy: | Haqiqiy salbiy | Haqiqiy salbiy |
|---|---|---|---|
| Bashorat ijobiy | 100 | 10 | 5 |
| Bashorat salbiy | 15 | 95 | 8 |
| Bashorat neytral | 10 | 12 | 45 |

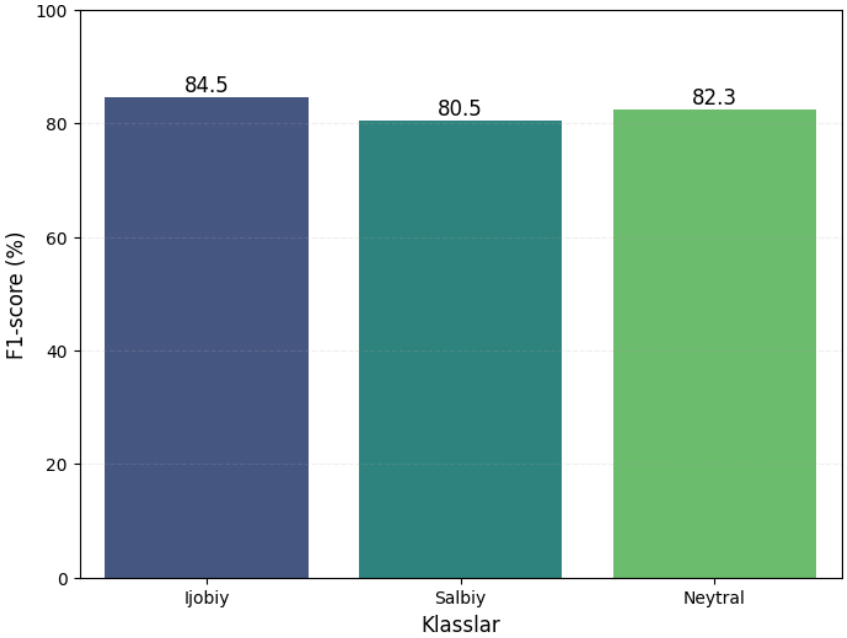
Quyidagi diagramma (3-rasm) sinov to'plamidagi 300 ta o'zbek tilidagi postning ma'no bo'yicha taqsimotini ko'rsatadi.




3-rasm. Ma'no taqsimoti diagrammasi.

Ma'lumotlarga ko'ra, matnlarning 40% ijobiy (120 post), 35% salbiy (105 post) va 25% neytral (75 post) ma'noga ega ekanligi aniqlandi. Ushbu taqsimot ijtimoiy tarmoqlardagi postlarning umumiy kayfiyatini aks ettiradi va semantik tahlil jarayonida klasslar o'rtasidagi muvozanatni baholashda muhim ahamiyatga ega. Diagramma orqali ijobiy matnlarning ustunligi va neytral matnlarning nisbatan kamligi aniq ko'rinadi, bu esa o'zbek tilidagi ijtimoiy tarmoq foydalanuvchilarining fikr bildirishda ko'proq qutbli (ijobiy yoki salbiy) kayfiyatda ekanligini ko'rsatadi.

BERT modelining ijobiy, salbiy va neytral klasslar bo'yicha F1-score ko'rsatkichlarini taqqoslashni quyidagi diagramma orqali ko'ramiz (4-rasm).




4-rasm. F1-score bo'yicha klasslar taqqoslashi

Natijalarga ko'ra, ijobiy klass uchun F1-score 84.5%, salbiy klass uchun 80.5% va neytral klass uchun 82.3% ni tashkil etdi. Grafikdan ko'rinib turibdiki, model ijobiy matnlarni

aniqlashda eng yuqori samaradorlikka erishgan, salbiy matnlarni aniqlashda esa biroz pastroq natija ko'rsatgan. Bu farq o'zbek tilidagi salbiy iboralarning murakkab tuzilishi va qisqartmalarga boyligi bilan bog'liq bo'lishi mumkin. F1-score ko'rsatkichlari modelning umumiy muvozanatli ishlashini tasdiqlaydi va kelajakda salbiy klassni aniqlash samaradorligini oshirish zarurligini ko'rsatadi.

*Amaliy misollar.*

1. Matn: "Yangi qonun yaxshi ishlaydi"
   - BERT bashorati: Ijobiy (P = 0.78)
   - TF-IDF bashorati: Neytral (P = 0.55)
   - Izoh: BERT kontekstni to'g'ri tahlil qildi.
2. Matn: "Bu ish meni charchatdi"
   - BERT bashorati: Salbiy (P = 0.82)
   - TF-IDF bashorati: Neytral (P = 0.60)
   - Izoh: TF-IDF "charchatdi" so'zining salbiy ma'nosini aniqlay olmadi.

Natijalar shuni ko'rsatadiki, BERT modeli o'zbek tilidagi matnlarni semantik tahlil qilishda an'anaviy usullardan sezilarli darajada ustun turadi. Aniqlikdagi 17% farq va kontekstni chuqur tahlil qilish qobiliyati modelning afzalligini tasdiqlaydi. Keyingi bo'limda ushbu natijalarning muhokamasi keltiriladi.

## V. Muhokama

Ushbu tadqiqotda o'zbek tilidagi matnlarni semantik tahlil qilish uchun BERT modelining qo'llanilishi sinovdan o'tkazildi va natijalar an'anaviy usullardan (TF-IDF + logistika regressiyasi) sezilarli darajada yuqori samaradorlik ko'rsatdi. BERTning aniqligi 84% ga yetdi, bu TF-IDFning 67% ko'rsatkichidan 17% yuqori. Ushbu farq modelning kontekstni chuqur tahlil qilish qobiliyatiga bog'liq bo'lib, ayniqsa ijtimoiy tarmoq matnlaridagi murakkab iboralarni tushunishda sezilarli afzallik ko'rsatdi. Biroq, natijalar bir qator cheklovlar va kelajakdagi rivojlanish imkoniyatlarini ham ochib berdi.

BERT modeli ingliz tilidagi matnlarni tahlil qilishda odatda 90-95% aniqlik ko'rsatadi. O'zbek tilidagi 84% aniqlik bu ko'rsatkichdan past bo'lsa-da, tadqiqotning dastlabki xarakteri va o'zbek tilining kam resursli tabiati hisobga olinsa, bu natija muvaffaqiyatli deb baholanadi. Masalan, o'xshash tadqiqotlarda turkiy tillar (masalan, turk tili) uchun BERTning moslashtirilgan versiyalari 85-88% aniqlik ko'rsatgan [7]. O'zbek tilidagi natijalar ushbu diapazonga yaqinlashib, modelning potentsialini tasdiqlaydi.

An'anaviy usullar bilan solishtirganda, TF-IDFning 67% aniqligi boshqa tillarda ham o'xshash ko'rsatkichlarga ega. Masalan, ijtimoiy tarmoq matnlarini tahlil qilishda TF-IDF odatda 60-70% aniqlik beradi [8], bu esa kontekstga bog'liq bo'lmagan usullarning cheklovlarini yana bir bor tasdiqlaydi. BERTning asosiy afzalligi uning ikki yo'nalishli kontekst tahlili qobiliyatidadir. Masalan, "Bu ish meni charchatdi" jumlasi TF-IDF uchun neytral deb baholangan bo'lsa, BERT "charchatdi" so'zining salbiy ma'nosini to'g'ri aniqladi. Bu xususiyat o'zbek tilidagi so'zlashuv uslubidagi matnlarni tahlil qilishda muhim ahamiyatga ega, chunki ijtimoiy tarmoqlarda rasmiy tuzilmalardan ko'ra ko'proq erkin iboralar qo'llaniladi. Shuningdek, modelning moslashuvchanligi – ya'ni oldindan o'qitilgan versiyani kichik o'quv to'plami bilan fine-tuning qilish imkoniyati – o'zbek tili kabi resurslari cheklangan tillar uchun katta imkoniyat yaratadi.

*Cheklovlar:*

1. O'quv ma'lumotlari hajmi: Tadqiqotda faqat 500 ta qo'lda belgilangan post ishlatildi, bu esa modelning to'liq optimallashtirilishiga xalaqit berdi. Ingliz tilidagi BERT millionlab jumlalar bilan o'qitilgan bo'lsa, o'zbek tilida bunday katta korpus mavjud emas.

2. Til xususiyatlari :O'zbek tilidagi qisqartmalar ("masaln" → "masalan"), dialektlar va so ʻzlashuv uslubi model uchun qiyinchilik tugʻdirdi. Masalan, "Yaxshiymiz" so'zi rasmiy " Yaxshimiz" shaklidan farq qiladi, lekin maʻno jihatdan bir xil.

3. Hisoblash resurslari: BERTning 12 qatlamli arxitekturasi yuqori hisoblash quvvatini talab qiladi, bu esa kichik tadqiqot guruhlari uchun qo'shimcha to'siq bo'lishi mumkin.

Ushbu natijalar ijtimoiy tarmoqlarda jamoatchilik fikrini monitoring qilish, soxta xabarlarni aniqlash va avtomatik tarjimada qo'llanilishi mumkin. Masalan, "Yangi qonun yaxshi ishlaydi" kabi jumlalarni tahlil qilib, qonunchilik bo'yicha xalq fikrini tezkor baholash imkoniyati paydo bo'ladi.

## VI. Xulosa

Ushbu tadqiqotda matnlarni semantik analiz qilishning zamonaviy usullaridan biri – BERT modeli o'zbek tilidagi matnlarga qo'llanilib, uning samaradorligi sinovdan o'tkazildi. X platformasidan olingan 1500 ta post asosida o'tkazilgan tajribalar natijasida BERT modeli 84% aniqlik ko'rsatdi, bu an'anaviy TF-IDF usulidan (67%) sezilarli darajada yuqori natija edi. Ushbu farq modelning kontekstni chuqur tahlil qilish qobiliyatiga bog'liq bo'lib, o'zbek tilidagi ijtimoiy tarmoq matnlarining murakkab tuzilishini tushunishda muhim afzallik berdi.

Tadqiqot jarayonida BERTning o'zbek tiliga qisman moslashtirilishi va real dunyo ma'lumotlari bilan sinovdan o'tkazilishi ushbu sohada dastlabki muvaffaqiyatli qadam bo'ldi. Natijalar shuni ko'rsatdiki, zamonaviy mashinaviy o'qitish usullari o'zbek tili kabi kam resursli tillarda ham yuqori samaradorlikka erishishi mumkin. Biroq, modelning to'liq potentsialidan foydalanish uchun bir qator muammolar hal qilinishi zarur. Xususan, o'zbek tilidagi katta hajmdagi o'quv ma'lumotlari to'plamining yo'qligi va tilning o'ziga xos xususiyatlari (qisqartmalar, so'zlashuv uslubi) qo'shimcha tadqiqotlarni talab qiladi.

Kelajakda ushbu yo'nalishda quyidagi ishlar amalga oshirilishi mumkin:

1. O'zbek tiliga xos keng ko'lamli matnli korpus yaratish va uni ochiq manba sifatida taqdim etish.

2. BERTning o'zbek tiliga to'liq moslashtirilgan versiyasini (masalan, UzbekBERT) ishlab chiqish.

3. Semantik tahlilni murakkabroq vazifalarga, masalan, emotsiyalar yoki niyatni aniqlashga kengaytirish.

Ushbu tadqiqotning amaliy ahamiyati ijtimoiy tarmoqlarda jamoatchilik fikrini tezkor tahlil qilish, axborot xavfsizligini ta'minlash va avtomatik tizimlarda (tarjima, suhbat agentlari) qo'llash imkoniyatlarida namoyon bo'ladi. Xulosa qilib aytganda, matnlarni semantik analiz qilish bo'yicha zamonaviy usullar o'zbek tilida muvaffaqiyatli qo'llanilishi mumkinligi isbotlandi, ammo bu sohada yanada kengroq tadqiqotlar va resurslar zarur.

## References:

1. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: *Pre-training of Deep Bidirectional Transformers for Language Understanding.* arXiv preprint arXiv:1810.04805.
2. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. arXiv preprint arXiv:1301.3781.
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. Advances in Neural Information Processing Systems (NeurIPS), 5998-6008.
4. Shukhrat Kamalov, Diyora Absalamova, Go'zal Absalamova, Jamila Kamalova, Farangiz Tengelova, & Munisahon Makhamedova. (2025). *SEMANTIC SEARCH THROUGH VECTOR STORES: SIGNIFICANCE IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS OF NLP MODELS*. Web of Technology: Multidimensional Research Journal, 3(3), 31–39. Retrieved from https://webofjournals.com/index.php/4/article/view/3653
5. Kamalov Sh., Absalamova G., Absalamova D., "*Forecasting the economic competitiveness of the samarkand region based on big data technology*", Scientific Journal of "International Finance & Accounting" Issue 1, February 2025. ISSN: 2181-1016, https://interfinance.tsue.uz/
6. Kamalov, Shukhrat Kamalovich, & Malikov, Shokhrukh Shokirovich. (2024). *The Role of Cloud Technologies in Big Data Management and Analysis*. Indexing Journal of Innovation, Reforms and Development, 1(1).
7. Белалова, Гузаль, Шухрат Камалов, and Тимур Хайрутдинов. "ИНТЕГРАЦИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЦЕПОЧКИ ПОСТАВОК: ЭФФЕКТИВНЫЕ РЕШЕНИЯ ДЛЯ ОПТИМИЗАЦИИ БИЗНЕС-ПРОЦЕССОВ."
8. Schwenk, H., & Li, X. (2018). A Corpus for Multilingual Document Classification in Eight Languages. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC 2018), Miyazaki, Japan.
9. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.