# BERT TIL MODELIDAN FOYDALANGAN HOLDA O‘ZBEK TILI LINGVISTIK MUAMMOLARINI YECHISH

**Jumanazarov Mardonbek Davronbek o‘g‘li**
jumanazarovmardonbek99@gmail.com,
UrDU magistranti
**Madatov Xabibulla Axmedovich**
fizika-matematika fanlari nomzodi
khabibulla@urdu.uz,
UrDU dotsenti

**Annotatsiya.** BERTbek nomi bilan ma’lum bo‘lgan o‘zbek BERT modelining paydo bo‘lishi o‘zbek tili uchun tabiiy tilni qayta ishlashda (NLP) muhim bosqichni ko‘rsatadi. Ushbu maqola BERTbekning turli lingvistik vazifalarni, xususan, hissiyotlarni tahlil qilish, matnlarni tasniflash, niqoblangan tilni modellashtirish (MLM) va nomuhim so‘zlarni olib tashlashdagi imkoniyatlarini o‘rganadi. Tadqiqotchilar va ishlab chiquvchilar ushbu oldindan o‘rgatilgan modeldan foydalanish orqali o‘zbek tilidagi NLP ilovalarining samaradorligi va aniqligini oshirishi, til texnologiyasi va lingvistik tadqiqotlardagi yutuqlarni qo‘llab-quvvatlashi mumkin.

**Abstract**. The advent of the Uzbek BERT model, known as BERTbek, marks a significant milestone in natural language processing (NLP) for the Uzbek language. This essay explores the capabilities of BERTbek in addressing various linguistic tasks, specifically sentiment analysis, text classification, masked language modeling (MLM), and stopwords removal. By leveraging this pre-trained model, researchers and developers can enhance the efficiency and accuracy of NLP applications in Uzbek, fostering advancements in language technology and linguistic research.

**Аннотация.** Появление узбекской модели BERT, известной как BERTbek, знаменует собой важную веху в области обработки естественного языка (NLP) для узбекского языка. В этом эссе исследуются возможности БЕРТбека в решении различных лингвистических задач, в частности, анализа настроений, классификации текста, моделирования языка в масках (MLM) и удаления стоп-слов. Используя эту предварительно обученную модель, исследователи и разработчики могут повысить эффективность и точность приложений НЛП на узбекском языке, способствуя развитию языковых технологий и лингвистических исследований.

**Kalit so‘zlar:** *NLP, BERT, o‘zbek tili, sentiment tahlil, matnlar tasnifi, MLM (maskalangan tilni modellashtirish), nomuhim so‘zlar.*

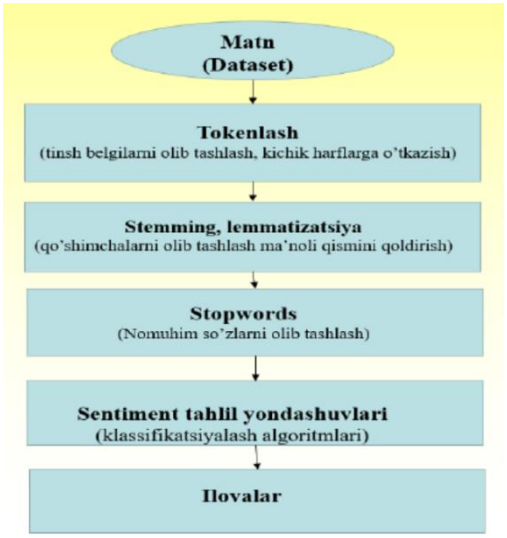
## Kirish

BERT (Transformerlardan ikki tomonlama kodlovchi vakilliklari) modellari matnning chuqur kontekstli tasvirlarini taqdim etish orqali NLP sohasida inqilob

qildi. O‘zbek tili uchun mo‘ljallangan BERTbek dasturining ishlab chiqilishi avval ixtisoslashtirilgan hisoblash resurslari yo‘qligi sababli qiyin bo‘lgan NLP vazifalari uchun yangi yo‘llarni ochadi. O‘zbekcha matnning katta jamlanmasida tayyorlangan ushbu model tilning nozik tomonlarini qamrab olgan bo‘lib, yanada nozik va samarali lingvistik tahlil qilish imkonini beradi. BERTbekning hissiyotlarni tahlil qilish, matn tasnifi, MLM va nomuhim so‘zlarni olib tashlashda qo‘llanilishi uning ko‘p qirraliligi va o‘zbek kontekstidagi lingvistik vazifalarni o‘zgartirish imkoniyatlarini ko‘rsatadi.

**NLP vazifalari**

Sentiment tahlil. BERTbek jamoatchilik fikrini tushunish, bozor tadqiqotlari va ijtimoiy media monitoringi uchun muhim bo‘lgan matn segmentlarining emotsional ohangini aniqlaydigan hissiyotlarni tahlil qilishda ustundir. BERTbek orqali kiritilgan matnni qayta ishlash orqali his-tuyg‘ularni ijobiy, salbiy yoki neytral deb aniq tasniflash mumkin, bu esa o‘zbekzabon jamoalarning jamoaviy kayfiyati yoki fikri haqida qimmatli ma’lumotlarni taqdim etish imkonini beradi.




O‘zbek tilidagi matnlarni sentiment tahlil qilish bosqichlari dastlab berilgan matnni fundamental NLP qadamlardan, masalan transliteratsiya [Kutlimuratova, Kuriyozov, Tillaeva, 2022: 161-171], tokenlash [Sharipov, Salayev, Matlatipov, 2021], hamda stopwrd(nomuhim so‘z)lardan tozalash [Kutlimuratova, 2021] kabi qadamlardan o‘tkazib, keyin sentiment tahlil qiluvchi modelga kerakli formatda yetkazib berishdan iborat. Bu bosqichlar quidagi blok-sxemada ko‘rsatilgan.

Lingvistik qoidalar qurish usulida salbiy va ijobiy ma‘no beruvchi so‘zlar bazasidan yoki umumiy lingvistik qoidalardan foydalaniladi. Sentimental gaplarda sifatlar ko‘p qo‘llanilganligi uchun sifatlarning orttirma darajasini yasovchi so‘zlarni ham kiritish orqali matnning salbiy yoki ijobiy ekanligini aniqroq aytishimiz mumkin.

Bu turdagi dasturga oddiy misol:

import re

```
pos_suzlar="yaxshi, a’lo, chiroyli, aqlli, go‘zal, yoqdi, bo‘ladi, ajoyib, ko‘p, qiziqarli"
neg_suzlar="yomon, rasvo, xunuk, aqlsiz, dabdala, qiziqarsiz, zerikarli, yoqmadi"
kuchaytiruvchi_suzlar="juda, eng, bag‘oyatda, g‘oyat, nihoyatda, behad "
gap=input("Matn kiriting: ")
gap1=gap.split()
t=0
for i in gap.split():
   if re.findall(i,pos_suzlar):
      t=1
      if re.findall(j, kuchaytiruvchi_suzlar):
         print("Bu ijobiy gapligi aniq")
         break
      print("Bu ijobiy gap")
      break
   elif re.findall(i, neg_suzlar):
      t=1
      if re.findall(j, kuchaytiruvchi_suzlar):
         print("Bu salbiy gapligi aniq")
         break
      print("Bu salbiy gap")
      break
   j=i
if t==0: print("Aniqlab bo‘lmadi.")
```

**Matnlar tasnifi**

Matnlarni tasniflashda BERTbek matnga oldindan belgilangan toifalarni tayinlaydi, bu esa kontentni tashkil etish, mavzuni ochish va ma’lumot qidirishni osonlashtiradi. Yangilik maqolalarini saralash, ilmiy maqolalarni tasniflash yoki veb-kontentni tartibga solish bo‘ladimi, BERTbekning chuqur o‘rganish imkoniyatlari tasniflash topshiriqlarida yuqori aniqlik va dolzarblikni ta’minlaydi, o‘zbek tilida ma’lumotlarga kirish va boshqarishni yaxshilaydi.

Niqoblangan tilni modellashtirish (MLM)

BERTning asosiy o‘quv komponenti bo‘lgan MLM jumlada yetishmayotgan so‘zlarni bashorat qilishni o‘z ichiga oladi. BERTbekning MLM tilidagi malakasi tilni ilg‘or tushunish va yaratish imkonini beradi, matnni to‘ldirish, avtomatik tuzatish va til o‘rgatish kabi vazifalarni bajarishda yordam beradi. Bu qobiliyat o‘zbek tilidan dinamik foydalanishni qo‘llab-quvvatlovchi tilga asoslangan intellektual dasturlarni ishlab chiqish uchun juda muhimdir.

Nomuhim so‘zlarni olib tashlash

BERT modellarining to‘g‘ridan-to‘g‘ri qo‘llanilishi bo‘lmasa-da, nomuhim so‘zlarni olib tashlash NLP-da dastlabki ishlov berishning muhim bosqichidir.

BERTbek kontekst va foydalanish asoslarini tushunish orqali nomuhim so‘zlar ro‘yxatini aniqlashtirishga yordam beradi va shu bilan quyi oqim NLP vazifalarining aniqligini oshiradi. BERTbek tegishli bo‘lmagan so‘zlarni aniqlash va yo‘q qilish orqali o‘zbek tilidagi lingvistik tahlillarning yo‘nalishi va samaradorligini oshiradi.

Xulosa

O‘zbek BERT modeli BERTbek o‘zbek tilidagi turli lingvistik vazifalarni hal qilish uchun yangi vosita hisoblanadi. Uning his-tuyg‘ularni tahlil qilish (sentiment tahlil), matn tasnifi, MLM va nomuhim so‘zlarni olib tashlashda qo‘llanilishi modelning moslashuvi va o‘zbekcha matnni qayta ishlash va tushunishdagi kuchini ko‘rsatadi. BERTbek nafaqat lingvistik tadqiqotlar va NLP ilovalarini ishlab chiqishni olg‘a siljitadi, balki o‘zbek tilining texnologik imkoniyatlarini kengaytirishga sezilarli hissa qo‘shadi va o‘zbek tilida so‘zlashuvchilar uchun tilni qayta ishlash yanada qulay, aniq va nozik bo‘lgan kelajakni va’da qiladi.

## Foydalanilgan adabiyotlar:

1. Salaev, U. I., Kuriyozov, E. R., & Matlatipov, G. R. (2023, November). Design and Implementation of a Tool for Extracting Uzbek Syllables. In *2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE)* (pp. 1750-1755). IEEE.

2. Madatov, K. (2019). A prolog format of uzbek WordNet’s entries. *Human Language Technology as a Challenge for Computer Science and Linguistics*, 316-320.

3. Allaberdiev, B., Matlatipov, G., Kuriyozov, E., &Rakhmonov, Z. (2024). Parallel texts dataset for Uzbek-Kazakh machine translation. *Data in Brief*, 110194.

4. Kutlimuratova, B., Kuriyozov, E., &Tillaeva, M. (2022). TEACHING ENGLISH AS A FOREIGN LANGUAGE FOR PRIMARY SCHOOL CHILDREN: LITERATURE REVEIW. *FOREIGN LANGUAGE TEACHING AND APPLIED LINGUISTICS*, 161-171.

5. Sharipov, M., Salaev, U., & Matlatipov, G. (2021). O‘ZBEK TILI FE’L SO‘Z TURKUMI UCHUN CHEKLI AVTOMATLAR ASOSIDA STEMMING ALGORITMINI YARATISH. *COMPUTER LINGUISTICS: PROBLEMS, SOLUTIONS, PROSPECTS*, *1*(1).

6. Kutlimuratova, B. (2021). Uzbek students learning English as a foreign language: Error analysis using corpora.

7. Sharipov, M., &Salaev, U. (2022). Uzbek affix finite state machine for stemming. *arXiv preprint arXiv:2205.10078*.

8. Madatov, K., Bekchanov, S., & Vičič, J. (2023). Uzbek text summarization based on TF-IDF. *arXiv preprint arXiv:2303.00461*.

9. Niyazmetova, Kumushoy & Kuriyozov, Elmurod. (2023). O‘zbek tilidagi matnlarni sentiment tahlil qilish.