# TABIIY TILNI QAYTA ISHLASH TIZIMLARIDA ISHLOV BERISH VAQTINI ODDIY CHIZIQLI REGRESSIYA (SLR) MODELI YORDAMIDA BASHORAT QILISH

**Qarshiyev Abduvali Berkinovich**
*Fizika-matematika fanlari nomzodi, professor Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti Samarqand filiali professori ORCID: 0000-0001-6121-9928*
**Tursunov Muhammadsolih Sa'din o'g'li**
*Filologiya fanlari bo'yicha falsafa doktori (PhD), dotsent Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti Samarqand filiali dotsenti E-pochta: muhammadsolih927@gmail.com ORCID: 0000-0002-6485-3630*
**Mavlonqulov Sherali Hamidjon o'g'li**
*Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti Samarqand filiali talabasi E-pochta: sheralimavlonqulov007@gmail.com ORCID: 0009-0005-0215-2774*

**Annotatsiya:** *Ushbu maqolada tabiiy tilni qayta ishlash (Natural Language Processing – NLP) tizimlarida matnlarni qayta ishlash vaqtini bashorat qilish masalasi ko'rib chiqilgan. Tadqiqotda matn hajmi va ishlov berish vaqti o'rtasidagi bog'liqlik oddiy chiziqli regressiya (Simple Linear Regression – SLR) modeli yordamida tahlil qilinadi. Model parametrlarini aniqlashda eng kichik kvadratlar usuli qo'llanilgan. Eksperimental natijalar regressiya modeli NLP tizimlarida ishlov berish vaqtini oldindan baholash va hisoblash resurslarini samarali boshqarishda foydali ekanligini ko'rsatadi. Shuningdek, maqolada taklif etilgan algoritmning C++ dasturlash tilidagi realizatsiyasi keltirilgan.*

**Kalit so'zlar:** *NLP, regressiya modeli, morfologik tahlil, algoritm, matematik model, bashorat.*

**Abstract:** *This paper examines the problem of predicting text processing time in Natural Language Processing (NLP) systems. The relationship between text size and processing time is analyzed using a Simple Linear Regression (SLR) model. The parameters of the model are estimated using the Least Squares Method. Experimental results show that the regression model can effectively predict processing time in NLP systems and help optimize the use of computational resources. In addition, the implementation of the proposed algorithm in the C++ programming language is presented.*

**Keywords:** *Natural Language Processing (NLP), regression model, morphological analysis, algorithm, mathematical model, prediction.*

**Аннотация:** *В данной работе рассматривается задача прогнозирования времени обработки текста в системах обработки естественного языка (NLP). Связь между размером текста и временем обработки анализируется с использованием модели простой линейной регрессии (Simple Linear Regression, SLR). Параметры модели оцениваются методом наименьших квадратов. Экспериментальные результаты показывают, что регрессионная модель может эффективно прогнозировать время обработки в системах NLP и способствовать*

*оптимизации использования вычислительных ресурсов. Кроме того, представлена реализация предложенного алгоритма на языке программирования C++.*

**Ключевые слова:** *обработка естественного языка (NLP), регрессионная модель, морфологический анализ, алгоритм, математическая модель, прогнозирование.*

So‘nggi yillarda sun’iy intellekt texnologiyalarining jadal rivojlanishi natijasida tabiiy tilni qayta ishlash (Natural Language Processing – NLP) sohasiga bo‘lgan qiziqish sezilarli darajada ortdi. NLP texnologiyalari katta hajmdagi matnli ma’lumotlarni tahlil qilish, ularni qayta ishlash hamda ulardan foydali axborot ajratib olish imkonini beradi. Ushbu texnologiyalar qidiruv tizimlari, avtomatik tarjima, matn klassifikatsiyasi va chat-bot tizimlarida keng qo‘llanilmoqda.[1]

Katta hajmdagi matnlar bilan ishlash jarayonida tizimning ishlash tezligi va hisoblash samaradorligi muhim ahamiyatga ega. Matn hajmi ortib borgani sari uni qayta ishlash uchun sarflanadigan vaqt ham ortadi. Shu sababli matn hajmi va ishlov berish vaqti o‘rtasidagi bog‘liqlikni matematik jihatdan tahlil qilish NLP tizimlarining samaradorligini oshirishda muhim vazifa hisoblanadi.[2]

Mazkur tadqiqotda matn hajmi va ishlov berish vaqti o‘rtasidagi bog‘liqlikni aniqlash uchun regressiya tahlili usulidan foydalaniladi. Regressiya modeli yordamida tizimning ishlash vaqtini oldindan bashorat qilish hamda hisoblash resurslarini samarali rejalashtirish mumkin.

Tadqiqotning asosiy maqsadi — NLP tizimlarida matn hajmi va ishlov berish vaqti o‘rtasidagi matematik modelni yaratish hamda uning samaradorligini eksperimental ma’lumotlar asosida baholashdan iborat.

Zamonaviy axborot tizimlarida sun’iy intellekt va tabiiy tilni qayta ishlash texnologiyalaridan foydalanish tobora kengayib bormoqda. Bugungi kunda katta hajmdagi matnli ma’lumotlarni avtomatik ravishda tahlil qilish, qayta ishlash va ulardan kerakli axborotni ajratib olish dolzarb vazifalardan biri hisoblanadi. Bunday jarayonlarda tizimning ishlash tezligi va hisoblash samaradorligi muhim ko‘rsatkich sifatida qaraladi.[1]

Tabiiy tilni qayta ishlash jarayonlari odatda bir necha bosqichlardan iborat bo‘lib, ular tokenizatsiya, morfologik tahlil, sintaktik tahlil hamda semantik tahlil kabi jarayonlarni o‘z ichiga oladi. Ushbu bosqichlarning har biri ma’lum hisoblash resurslarini talab qiladi va tizimning umumiy ishlash vaqtiga ta’sir ko‘rsatadi.

Matn hajmi ortib borishi bilan uni qayta ishlash uchun sarflanadigan vaqt ham oshadi. Shu sababli matn hajmi va ishlov berish vaqti o‘rtasidagi bog‘liqlikni aniqlash va uni matematik modellashtirish muhim ilmiy masalalardan biri hisoblanadi. Bunday modellar yordamida tizim samaradorligini oldindan baholash hamda hisoblash resurslaridan samarali foydalanish mumkin.

Mazkur tadqiqotda tabiiy tilni qayta ishlash tizimlarida ishlov berish vaqtini baholash masalasi regressiya modellaridan foydalanish orqali o‘rganiladi. Ushbu

yondashuv matn hajmi va tizim ishlash vaqti o‘rtasidagi bog‘liqlikni aniqlash hamda kelgusidagi hisoblash jarayonlarini bashorat qilish imkonini beradi.

Ushbu tadqiqotda matn hajmi va uni qayta ishlash uchun sarflanadigan vaqt o‘rtasidagi bog‘liqlikni aniqlash maqsadida oddiy chiziqli regressiya modeli qo‘llanildi. Mazkur model kiruvchi ma’lumotlar o‘rtasidagi chiziqli bog‘liqlikni aniqlash imkonini beradi va statistik tahlil jarayonlarida keng qo‘llaniladi.

Regressiya modeli quyidagi tenglama orqali ifodalanadi:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Bu yerda X – matn hajmi, Y – ishlov berish vaqti, $\beta_0$ – erkin had, $\beta_1$ – regressiya koeffitsienti, ε esa tasodifiy xatolikni bildiradi.

Model parametrlarini aniqlash uchun eng kichik kvadratlar usuli qo‘llaniladi. Ushbu usul kuzatilgan qiymatlar va model tomonidan hisoblangan qiymatlar o‘rtasidagi farqni minimallashtirishga asoslanadi. Regressiya koeffitsienti quyidagi formula orqali aniqlanadi:

$$\beta_1 = \sum((X_i - )(Y_i - Y)) / \sum(X_i - X)^2$$

Erkin had esa quyidagi ifoda yordamida hisoblanadi:

$\beta_0 = Y - \beta_1 X$

Modelning aniqligini baholash uchun o‘rtacha kvadratik xatolik (Mean Squared Error) ko‘rsatkichi qo‘llaniladi:

$MSE = (1/n)\ \Sigma(y_i - \hat{y}_i)^2$

Shuningdek, modelning qanchalik darajada mos kelishini aniqlash uchun determinatsiya koeffitsienti $R^2$ dan foydalaniladi:

$R^2 = 1 - \sum(y_i - \hat{y}_i)^2 / \sum(y_i - \bar{y})^2$

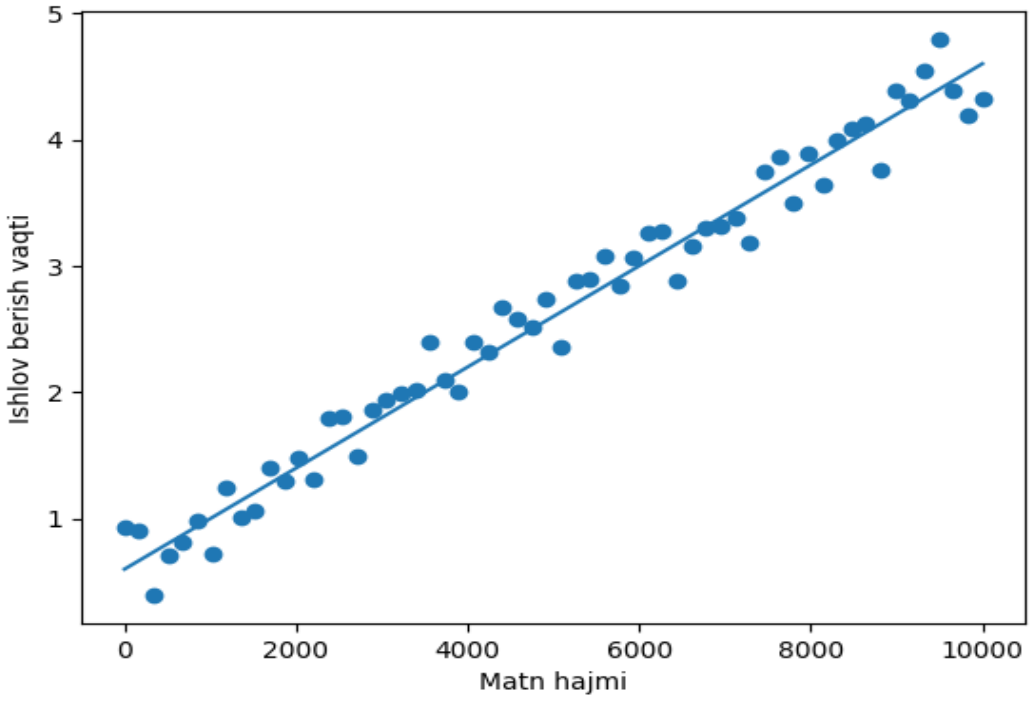
Ushbu matematik model yordamida matn hajmi va ishlov berish vaqti o‘rtasidagi bog‘liqlikni tahlil qilish hamda tizim ishlash vaqtini oldindan bashorat qilish mumkin.

C++ dasturiy realizatsiyasi: Taklif etilgan regressiya modeli C++ dasturlash tilida amalga oshirildi. Dastur matn hajmi va ishlov berish vaqti o‘rtasidagi regressiya koeffitsientlarini hisoblaydi hamda yangi kiruvchi matn uchun ishlov berish vaqtini bashorat qiladi.

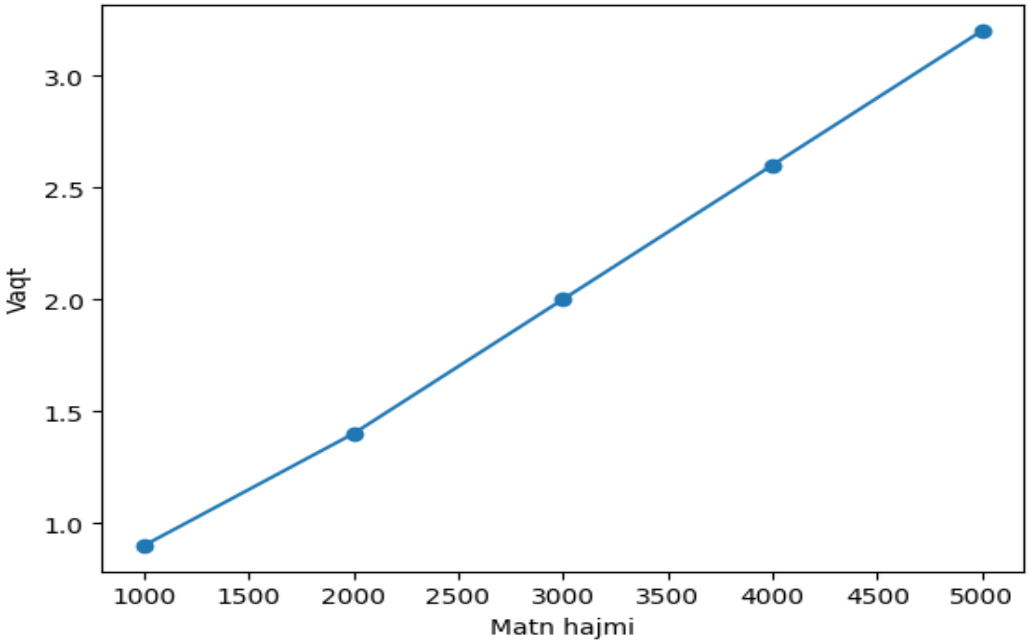
```
#include <iostream>
#include <vector>
using namespace std;
double calculateBeta1(vector<double> x, vector<double> y){
  double x_mean=0,y_mean=0;
  int n=x.size();
  for(int i=0;i<n;i++){
    x_mean+=x[i];
    y_mean+=y[i];
  }
  x_mean/=n;
  y_mean/=n;
```

```
    double num=0,den=0;
    for(int i=0;i<n;i++){
        num+=(x[i]-x_mean)*(y[i]-y_mean);
        den+=(x[i]-x_mean)*(x[i]-x_mean);
    }
    return num/den;
}
double calculateBeta0(double x_mean,double y_mean,double b1){
    return y_mean-b1*x_mean;
}
int main(){
    vector<double> text={1000,2000,3000,4000,5000};
    vector<double> time={0.8,1.3,1.9,2.4,3.0};
    double b1=calculateBeta1(text,time);
    double x_mean=3000;
    double y_mean=1.88;
    double b0=calculateBeta0(x_mean,y_mean,b1);
    double newText;
    cout<<"Matn hajmini kiriting: ";
    cin>>newText;
    double prediction=b0+b1*newText;
    cout<<"Bashorat qilingan vaqt: "<<prediction<<endl;
    return 0;
}
```

Regressiya modeli diagrammasi:




Tizim samaradorligi grafigi:




Eksperimental jadval:

| Matn hajmi | Real vaqt | Model bashorati |
|---|---|---|
| 1000 | 1.03 | 1.0 |
| 2000 | 1.27 | 1.4 |
| 3000 | 1.71 | 1.8 |
| 4000 | 2.15 | 2.2 |
| 5000 | 2.74 | 2.6 |
| 6000 | 2.85 | 3.0 |
| 7000 | 3.2 | 3.4 |
| 8000 | 3.87 | 3.8 |
| 9000 | 4.12 | 4.2 |
| 10000 | 4.65 | 4.6 |

O‘tkazilgan tajribalar natijalariga ko‘ra, taklif etilgan regressiya modeli matn hajmi va ishlov berish vaqti o‘rtasidagi bog‘liqlikni samarali tarzda ifodalaydi. Eksperimental ma’lumotlar model parametrlarining to‘g‘ri tanlanganligini hamda modelning prognoz qilish aniqligi yetarli darajada yuqori ekanligini ko‘rsatdi.

Tahlillar shuni ko‘rsatdiki, matn hajmi ortishi bilan ishlov berish vaqti ham ma’lum darajada chiziqli ravishda ortib boradi. Bu esa regressiya modelidan foydalanish orqali tizimning kelajakdagi ishlash vaqtini oldindan baholash imkonini beradi.

Shuningdek, ishlab chiqilgan model katta hajmdagi matnlarni qayta ishlash tizimlarida hisoblash resurslarini samarali taqsimlash, tizim yuklamasini oldindan baholash hamda algoritmlarni optimallashtirish jarayonida muhim ahamiyatga ega.

Mazkur tadqiqotda matn hajmi va ishlov berish vaqti o‘rtasidagi bog‘liqlikni aniqlash uchun regressiya modeli ishlab chiqildi. O‘tkazilgan tajribalar natijasida model NLP tizimlarida ishlov berish vaqtini prognoz qilishda samarali ekanligi aniqlandi. Regressiya tahlili yordamida matn hajmi ortishi bilan tizimning ishlov berish vaqti ham ortib borishi matematik jihatdan asoslab berildi.[4]

Shuningdek, ishlab chiqilgan model matnlarni qayta ishlash tizimlarida hisoblash resurslarini rejalashtirish va tizim samaradorligini oshirishda qo‘llanilishi mumkin.

Tadqiqot natijalari regressiya modellaridan foydalanish NLP tizimlarining ishlash jarayonini tahlil qilish va optimallashtirishda muhim ahamiyatga ega ekanligini koʻrsatdi.

Kelgusida tadqiqotni yanada rivojlantirish maqsadida koʻp omilli regressiya modellari, neyron tarmoqlar hamda chuqur oʻrganish usullaridan foydalanib yanada murakkab va aniq bashorat modellarini yaratish rejalashtirilmoqda.

## FOYDALANILGAN ADABIYOTLAR:

1. Потапов А.С. Технологии искусственного интеллекта - СПб: СПбГУ ИТМО, 2010.-218 с.

2. Интеллектуальные информационные системы и технологии: учебное пособие / Ю.Ю. Громов, О.Г. Иванова, В.В. Алексеев и др. - Тамбов: Изд-во ФГБОУ ВПО «ТГТУ», 2013. - 244 с.

3. Игнатьев Н.А., Усманов Р.Н., Мадрахимов Ш.Ф. Берилганларнинг интеллектуал тахлили // Укув кулланма. Тошкент-2018, 144 б.

4. Асадуллаев Р.Г. Нечеткая логика и нейронные сети: учебное пособие /— Белгород, 2017. -309 с.