ISNN 3030-3907

2026/4
VOLUME 2

DEVELOPMENT OF SCIENCE

ILMIY JURNAL



Development of science

Ilmiy jurnal

2026/4 VOLUME 2

ISSN 3030 -3907

Ilmiy jurnal OAK rayosatining 2024 yil 27 sentabrdagi 361-son qaroriga asosan 02.00.00 kimyo fanlari hamda 2025 yil 12 fevral 367- son qaroriga asosan 05.00.00 texnika fanlari bo‘yicha falsafa doktori (PhD) va fan doktori (DSc) ilmiy darajasiga talabgorlarning dissertatsiya ishlari yuzasidan dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan milliy ilmiy nashrlar ro‘yxatiga kiritilgan.

BOSH MUHARRIR:
Temirov Alisher Hoshim o‘g‘li – texnika fanlari falsafa doktori (PhD),Dotsent.
TAHRIRIYAT KENGASH RAISI:
Fozilov Sadriddin Fayzullayevich – texnika fanlari doktori, professor
TAHRIRIYAT KENGASH A’ZOLARI

TEXNIKA FANLARI/ TECHNICAL SCIENCES/ ТЕХНИЧЕСКИЕ НАУКИ

Bozorov G‘ayrat Rashidovich
Buxoro Davlat Texnika Universiteti, professor
Abdurahmonov Olim Rustamovich
Buxoro Davlat Texnika Universiteti, professor
Adizov Bobur Zamirovich
O‘zbekiston Respublikasi Fanlar akademiyasi Umumiy va noorganik kimyo instituti, professor
Hayitov Ruslan Rustamjonovich
Buxoro Davlat Texnika Universiteti, professor
Jumayev Qayum Karimovich
Buxoro Davlat Texnika Universiteti, professor.
Berdiyev Dorob Murodovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor
Maxmudov Rafiq Amonovich
Buxoro Davlat Texnika Universiteti, professor.
Tilloyev Lochin Ismatilloyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Raxmonov Qahramon Sanakulovich
Buxoro Davlat Texnika Universiteti, dotsent.
Barakayev Nusratullo Rajabovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor.
Jo‘rayev Xayrullo Fayziyevich
Buxoro Davlat Texnika Universiteti, professor.
Panoyev Erali Rajabboyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Ataullayev Sherzod Nabiullyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Malik Negmurodovich
Buxoro Davlat Texnika Universiteti, dotsent.
Axmedova Ozoda Bahronovna
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti.
Safarov Bahri Jumayevich
Buxoro Davlat Texnika Universiteti, dotsent.
Xaydarov Axtam Amonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Safarov Jasur Alijon o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Асадова Раъно Дилмуратовна
Тошкент Давлат техника университети доценти

KIMYO FANLARI/ CHEMICAL SCIENCES/ ХИМИЧЕСКИЕ НАУКИ

Axmedov Voxid Nizomovich
Buxoro Davlat Texnika Universiteti, professor
Umarov Baqo Bafoyevich
Buxoro Davlat Universiteti, professor
Kodirov Abduaxad Abduraximovich
Qarshi davlat Universiteti, professor
Qurbonov Mingniqul Jumagulovich
Qarshi davlat Universiteti, dotsent
Turemuratov Sharibay Naurizbaevich
O‘zbekiston Respublikasi Fanlar akademiyasi Qoraqalpog‘iston bo‘limi Bosh ilmiy kotibi, professor
Tursunov Kaxor Shonazarovich
Qarshi davlat Universiteti, professor
Do‘stov Hamro
Buxoro Davlat Texnika Universiteti, professor
Panoyev Nodir Shavkatovich
Buxoro Davlat Texnika Universiteti, dotsent.
Olimov Bobur Bahodir o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Kasimov Sherzod Abduazimovich
Termiz Davlat Universiteti, professor.
Aliqulov Rustam Valiyevich
Termiz Davlat Universiteti, professor
Tropov Xamza Tursunovich
Samarqand Davlat Universiteti, professor

IQTISODIYOT FANLARI/ ECONOMICS/ ЭКОНОМИЧЕСКИЕ НАУКИ

Nizamov Aslitdin Badreddinovich
Buxoro Davlat Texnika Universiteti, professor.
Boboyev Akmal Choriyevich
Buxoro Davlat Texnika Universiteti, professor.
Ergashxojayeva Shaxnoza Djasurovna
Toshkent davlat iqtisodiyot universiteti, professor
Saidaxmedova Nodira Ilxomovna
Toshkent davlat iqtisodiyot universiteti, professor
Muxsinov Bekzod Toxirovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shakarov Qulmat Ashirovich
Renessans ta'lim universiteti,dotsent.
Jumaeva Zulfiya Qayumovna
Buxoro Davlat Texnika Universiteti, dotsent.
Norova Salomat Yusupovna
Buxoro Davlat Texnika Universiteti, dotsent.
Avezova Shaxnoza Maxmudjonovna
Buxoro Davlat Texnika Universiteti, dotsent.
Jalilov Jamshid G‘anijonovich
Toshkent davlat iqtisodiyot universiteti, dotsent
Xasanova Gulruh Djumanazarovna
Buxoro Davlat Texnika Universiteti, dotsent.

PEDAGOGIKA FANLARI/ PEDAGOGICAL SCIENCES / ПЕДАГОГИЧЕСКИЕ НАУКИ

Siddiqova Sadoqat G‘afforovna
Buxoro Davlat Texnika Universiteti, dotsent.
Zoyirov Erkin Xalilovich
Buxoro Davlat Texnika Universiteti, DSc, professor
Gaffarov Feruz Hasanovich
Puchon Universiteti, *professor*
Ruziyev Davron Yuldashevich
Buxoro Davlat Universiteti, professor
Muradova Firuza Rashidovna
Buxoro Davlat Texnika Universiteti, professor.
Yunusova Gulandon Samiyevna
Buxoro Davlat Texnika Universiteti, professor.
To‘xtayeva Zebo Sharifovna
Buxoro Davlat Texnika Universiteti, professor.
Navro‘zova Gulchixra Nigmatovna
Buxoro Davlat Texnika Universiteti, dotsent.
Uzoqova Laylo Polvonovna
Buxoro muhandislik – texnologiya instituti, dotsent.
To‘raqulova Marjona Qiyom qizi
Buxoro Davlat Texnika Universiteti, dotsent.
Muradova Zarina Rashidovna
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Sanjar Aslonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shodiyev Jahongir Jo‘raqulovich
Buxoro Davlat Texnika Universiteti, dotsent
Gaffarov Laziz Hasanovich
Buxoro muhandislik – texnologiya instituti, dotsent.

Tahririyat a’zolarning to‘liq ro‘yxati jurnalning www.devos.uz saytida keltirilgan.

Tahririyat hay’ati a’zolarining tarkibini o‘zgartirish hamda qo‘shimcha kiritish yoki tahririyat a’zoligidan chiqarish bosh muharrir tomonidan amalga oshiriladi.

ISSN 3030-3907

Development of science ilmiy jurnali 2024 yil 15 aprel kuni №256445 sonli guvohnoma bilan Davlat ro‘yxatidan o‘tgan.

To‘plam elektron ko‘rinishda (pdf formatida) mualliflarga taqdim etiladi. To‘plamga kiritiladigan maqolalarning mazmuni, ulardagi statistik ma’lumotlar va me’yoriy hujjatlarning to‘g‘riligiga hamda har qanday fikr - mulohazalarga mualliflarning o‘zlari mas’ul hisoblanadilar. Talablarga javob bermaydigan maqolalar to‘plamga kiritilmaydi. Tashkiliy qo‘mita maqola matnini qisqartirish, qisman tuzatish kiritish, seksiyalarga joylashtirish huquqiga ega.

ISSN 3030-3907

CNN VA GRU ASOSIDA MATN HUJJATLARINI AVTOMATIK TASNIFLASH

Muhamediyeva D.T.
«Toshkent irrigatsiya va qishloq xo‘jaligini mexanizatsiyalash muhandislari instituti» milliy tadqiqot universiteti, Namangan davlat universiteti.
E-mail: dilnoz134@rambler.ru
Mamatov A.A.
«Toshkent irrigatsiya va qishloq xo‘jaligini mexanizatsiyalash muhandislari instituti» milliy tadqiqot universiteti, Namangan davlat universiteti.
E-mail:abduvali_mamatov@mail.ru

Annotatsiya. Ushbu tadqiqotda matn hujjatlarini tasniflash uchun avtomatik tizim ishlab chiqildi. Tizim konvolyutsion neyron tarmoqlari va eslab qoluvchi tarmoqlarni birlashtirib, shuningdek, so‘zlarning vaznini hisobga oluvchi mashina o‘rganishi metodlarini qo‘llaydi. Tajribalar AG News ma’lumotlar to‘plami asosida o‘tkazilib, tizim 120,000 hujjat bo‘yicha o‘qitildi va 7,600 hujjat bo‘yicha sinovdan o‘tkazildi. Natijalar shuni ko‘rsatdiki, taklif etilgan yondashuv an’anaviy usullarga nisbatan ancha yuqori aniqlikka ega bo‘ldi.

Kalit so‘zlar. Matn tasniflash, konvolyutsion neyron tarmoqlari, ksda saqlovchi tarmoqlar, so‘z vazni, mashina o‘rganishi.

Аннотация. В трех исследованиях была разработана автоматическая система классификации текста. Система использует методы машинного обучения для объединения сверточных нейронных сетей и сетей полноты для сбора весов слов. Эксперименты проводились на наборе данных AG News, где система была обучена на 120 000 документах и протестирована на 7600 документах. Результаты показывают, что предложенный метод достигает значительно более высокой точности, чем традиционный метод.

Ключевые слова: Классификация текста, сверточные нейронные сети, весовые коэффициенты слов, машинное обучение.

Abstract. In three studies, an automatic system for text classification was developed. The system uses machine learning techniques to combine convolutional neural networks and recall networks to collect word weights. The experiments were conducted on the AG News dataset, where the system was trained on 120,000 documents and tested on 7,600 documents. The results show that the proposed method achieves significantly higher accuracy than the traditional method.

Keywords: Text classification, convolutional neural networks, convolutional neural networks, word weight, machine learning.

KIRISH.

ISSN 3030-3907

So‘nggi yillarda turli sohalarda matn hujjatlari soni keskin oshdi. Hujjatlarni avtomatik tasniflash muhim masala bo‘lib, soxta yangiliklarni aniqlash, mijoz so‘rovlarini toifalash, hisobotlarni tahlil qilish kabi amaliyotlarda qo‘llaniladi. Matnlarni dasturiy yo‘l bilan tahlil qilish murakkab bo‘lib, tabiiy tilning noaniqligi va hujjatlarning turlicha tuzilishi sababli klassik usullar yetarli natija bermaydi. Shu sababli ushbu ishda konvolyutsion va eslab qoluvchi neyron tarmoqlar, shuningdek, so‘z vazni asosida mashina o‘rganishi metodlari birlashtirilgan tizim ishlab chiqildi. Bu yondashuv matnlarni to‘g‘ri tasniflash va real amaliyotlarda qo‘llash imkonini beradi [1-2].

Avtomatik matn hujjatlarini tasniflash sun’iy intellekt va tabiiy tilni qayta ishlash (NLP) sohalaridagi jadal rivojlanish natijasida muhim tadqiqot yo‘nalishiga aylandi. Tadqiqotlarning aksariyati mashinaviy o‘rganish va chuqur o‘rganishga asoslangan modellar yordamida turli soha va tillarga moslashtirilgan tasniflash yechimlarini taklif etishga qaratilgan. Dastlabki tadqiqotlarda an’anaviy mashinaviy o‘rganish modellaridan foydalanilgan bo‘lsa, keyingi ishlar chuqur neyron tarmoqlarning yuqori darajadagi semantik xususiyatlarni avtomatik o‘rganishdagi ustunligini namoyish etdi. Masalan, Anh va boshqalar DNN modellarining past darajadagi xususiyatlardan murakkab tushunchalarni aniqlashda samaraliroq ekanini ko‘rsatgan. Shunga o‘xshash tarzda, Katrina va hamkorlari arab tilidagi matnlar uchun chuqur o‘rganishga asoslangan modellar an’anaviy regressiya usullariga nisbatan yuqori aniqlikni ta’minlashini isbotladilar. Bu natijalar chuqur o‘rganish yondashuvlarining tilga xos vazifalarda muhim afzalliklarga ega ekanini ko‘rsatadi [2-4].

Soha-spetsifik ilovalarda chuqur o‘rganish usullarining samaradorligi hujjat tasvirlari va murakkab ma’lumotlar bilan ishlashda ham tasdiqlandi. Chen va boshqalar Faster R-CNN modelini qo‘llab, energiya boshqaruv tizimlaridagi murakkab hujjatlarni avtomatik tasniflashga erishgan. Shuningdek, NLP texnikalaridan foydalanilgan tibbiy ilovalarda, xususan, shizofreniyani erta aniqlashda, matn asosidagi modellar diagnostika jarayonini qo‘llab-quvvatlash imkoniyatiga ega ekanligi ko‘rsatildi [5-7].

Bir qator tadqiqotlar tuzilmagan ma’lumotlar bilan ishlashda xususiyatlarni ajratib olish va tasvirlash muammosiga e’tibor qaratdi. BERT kabi kontekstual embedding modellari matnning ikki yo‘nalishli semantikasini hisobga olib, tasniflash aniqligini sezilarli darajada oshirgan. Shu bilan birga, ontologiyaga asoslangan modellar va chuqur o‘rganishning kombinatsiyasi murakkab semantik tuzilmalarni aniqlashda samarali ekanini ko‘rsatdi. Biroq, ushbu modellar ko‘pincha yuqori hisoblash resurslarini talab qilishi bilan cheklanadi [8-10].

ISSN 3030-3907

Nomutanosib ma’lumotlar muammosini hal etishga qaratilgan ishlarda nazoratli va nazoratsiz o‘rganish yondashuvlari birlashtirilib, tasniflash natijalarining barqarorligi oshirilgan. Shuningdek, IGFSS, VGFSS va OTFS kabi xususiyat tanlash algoritmlari sinflar o‘rtasida muvozanatli termin taqsimotini ta’minlab, klassifikatsiya samaradorligini yaxshilagan. Biroq, ushbu usullarning aksariyati muayyan ma’lumotlar to‘plamlariga moslashtirilgan bo‘lib, umumiylashuv qobiliyati cheklangan [11-12].

Resurslar cheklangan muhitlarda olib borilgan tadqiqotlar chuqur o‘rganish modellarining hisoblash murakkabligini kamaytirishga qaratilgan. Past resursli tillar uchun ishlab chiqilgan yengillashtirilgan DNN arxitekturalari va few-shot yondashuvlar trening vaqtini qisqartirishga xizmat qilgan. Shunga qaramay, bu usullar ko‘pincha kontekstni yetarlicha chuqur ifodalay olmaslik yoki turli domenlarga moslashishda qiyinchiliklarga duch keladi. Umuman olganda, mavjud adabiyotlar chuqur o‘rganish asosidagi modellar matn tasniflashda yuqori samaradorlikka ega ekanini ko‘rsatadi. Biroq, ko‘plab yondashuvlar ma’lum bir soha yoki tilga bog‘langan, katta hajmdagi o‘quv ma’lumotlarini talab qiladi hamda hisoblash resurslariga yuqori talab qo‘yadi. Shu bois, moslashuvchan, samarali va turli hujjat turlariga tatbiq etiladigan universal freymvorklar yetarli darajada o‘rganilmagan [13-15].

Mazkur tadqiqotda taklif etilayotgan Learning-based Text Classification (LbTC) algoritmiga asoslangan Avtomatik Matn Tasniflash Freymvorki (ATCF) ushbu bo‘shliqni to‘ldirishga qaratilgan bo‘lib, sayoz va chuqur o‘rganish modellarini yagona tizimda birlashtirish orqali turli hujjat turlari va amaliy sharoitlarda yuqori aniqlik va moslashuvchanlikni ta’minlashni maqsad qiladi. Ushbu maqolaning asosiy maqsadi matn hujjatlarini avtomatik tasniflash uchun samarali tizim ishlab chiqishdir. Bu tizim turli sohalardagi hujjatlarni, masalan, yangilik maqolalari, ilmiy hisobotlar yoki tijorat materiallarini toifalash imkonini beradi. Maqolaning vazifalari quyidagilardan iborat:

Matnlarni oldindan tozalash va ularni mashina o‘rganishi uchun tayyorlash metodlarini ishlab chiqish. So‘z vazni va ularning kontekstini hisobga oluvchi xususiyatlarni chiqarish. Konvolyutsion va eslab qoluvchi neyron tarmoqlarni birlashtirib, so‘z vazni asosidagi mashina o‘rganishi yondashuvi bilan samarali model yaratish. Tizimni sinovdan o‘tkazib, aniqlik, sezgirlik va umumiy samaradorlikni baholash.

Ushbu ishning ilmiy yangiligi bir necha jihatdan namoyon bo‘ladi. Birinchidan, matn tasniflashda an’anaviy modellardan farqli o‘laroq, konvolyutsion va eslab qoluvchi neyron tarmoqlarni birlashtirish orqali hujjatlarni yanada chuqurroq o‘rganish imkoniyati yaratilgan. Ikkinchidan, so‘zlarning

ISSN 3030-3907

vaznini hisobga oluvchi mashina o‘rganishi metodlari bilan integratsiya qilinib, modelning aniqligi sezilarli darajada oshirilgan. Uchinchi jihat – tizimni katta hajmdagi ma’lumotlar bilan sinovdan o‘tkazish va turli kategoriyalardagi hujjatlarda barqaror ishlashini ko‘rsatish, bu esa uni ilmiy va amaliy jihatdan foydali qiladi. Shu bilan birga, tizim neyron tarmoqlar va matn xususiyatlarini birlashtirish orqali yangi metodologik yondashuv sifatida tavsiflanadi.

Taklif etilgan tizim real hayotdagi bir nechta vazifalarni samarali bajarish imkonini beradi. Masalan, yangiliklar saytlari uchun avtomatik maqola tasniflash, ijtimoiy tarmoqlardagi postlarni mavzuga qarab saralash, ilmiy maqolalarni soha bo‘yicha kategoriyalash va tijorat hisobotlarini tezkor tahlil qilish mumkin. Shu bilan birga, tizim korporativ va davlat tashkilotlarida katta hajmdagi hujjatlarni avtomatik tartibga solish va boshqarish uchun qulay vosita bo‘lib xizmat qiladi. Barcha bu jihatlar tizimning amaliy ahamiyatini oshiradi va real hayotdagi vazifalarda qo‘llanilishini kafolatlaydi.

TADQIQOT METODOLOGIYASI.

Ushbu bo‘limda matn hujjatlarini avtomatik tasniflash uchun ishlab chiqilgan tizimning metodologiyasi batafsil tushuntiriladi. Tadqiqot uch asosiy bosqichdan iborat: matnni oldindan tozalash va xususiyatlarni chiqarish, klassik mashina o‘rganishi metodlari bilan tasniflash, hamda chuqur neyron tarmoqlarni TF-IDF bilan integratsiya qilish.

Matnlarni tahlil qilishdan oldin quyidagi jarayonlar amalga oshiriladi:HTML teglarni olib tashlash; Maxsus belgilarni chiqarish; Ortiqcha bo‘shliqlarni bartaraf etish; So‘zlarni kichik harfga o‘tkazish.

Matematik ko‘rinishda bu jarayon quyidagicha ifodalanadi:

$$x_i' = f_{\text{tozalash}}(x_i),$$

bu yerda x_i — kiruvchi matn, x_i' — tozalangan matn, f_{tozalash} — yuqoridagi jarayonlar funksiyasi.

Matnlarni sonli ko‘rinishga keltirish uchun so‘z vazni asosida TF--IDF usuli qo‘llaniladi:

$$\text{tf-idf}_{t,d} = \text{tf}_{t,d} \cdot \log \frac{N}{df_t},$$

bu yerda:

t — so‘z

d — hujjat

$\text{tf}_{t,d}$ — so‘zing hujjatdagi chastotasi

N — hujjatlar soni

df_t — so‘z qatnashgan hujjatlar soni

ISSN 3030-3907

Shundan so‘ng hujjatni vektorga aylantiramiz:

$$X_{\text{tfidf}} = [\text{tf-idf}_{1,d}, \ldots, \text{tf-idf}_{M,d}],$$

bu yerda M — eng ko‘p ishlatiladigan so‘zlar soni.

Logistic Regression

Har bir hujjatning k — chi toifaga tegishlilik ehtimoli:

$$P(y = k \mid X) = \frac{e^{\beta_k^T X}}{\sum_{j=1}^{K} e^{\beta_j^T X}}.$$

Model parametrlarini maksimal ehtimollik yordamida o‘rganamiz:

$$\hat{\beta} = \arg\max_{\beta} \prod_{i=1}^{N} P(y_i \mid X_i).$$

Random Forest

Random Forest bir nechta qaror daraxtlari orqali bashorat beradi:

$$h_t(X) = \arg\max_k \; p_t(y = k \mid X).$$

Oxirgi qaror esa daraxtlarning ovoziga asoslanadi:

$$\hat{y} = \text{mode}\{h_1(X), \ldots, h_T(X)\},$$

bu yerda T — daraxtlar soni.

Neyron tarmoqlar va xususiyatlarni birlashtirish jarayonida Embedding qatlamida

so‘zlar ketma-ketligini indekslarga aylantiramiz:

$$\text{seq}_i = [w_1, \ldots, w_L].$$

Shundan so‘ng so‘zlarni zich vektorlar bilan ifodalaymiz:

$$e_i = E[w_i], \qquad E \in R^{V \times d}.$$

bu yerda V — so‘zlar lug‘ati hajmi, d — embedding o‘lchami.

Konvolyutsion qatlam (CNN) da hujjatning mahalliy xususiyatlarini aniqlash:

$$c_i = f\left(\sum_{j=0}^{k-1} w_j \cdot e_{i+j} + b\right).$$

Global maksimal havola:

$$y_{\text{cnn}} = \max_i c_i.$$

Esda saqlovchi qatlam (GRU):

$$r_t = \sigma(W_r[h_{t-1}, e_t]),$$

$$z_t = \sigma(W_z[h_{t-1}, e_t]),$$

$$\tilde{h} = \tanh(W_h[r_t \odot h_{t-1}, e_t]),$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t.$$

ISSN 3030-3907

TF--IDF va neyron tarmoqlar integratsiyasi jarayonida CNN va GRU xususiyatlari TF--IDF vektori bilan birlashtiriladi:

$$f_{\text{fusion}} = \left[y_{\text{cnn}} \oplus h_T \oplus X_{\text{tfidf}} \right].$$

Chiqish qatlamida yumshoq maksimal (softmax) funksiyasi ishlatiladi:

$$\hat{y} = \text{softmax}(Wf_{\text{fusion}} + b).$$

Katergoriyali entropiya yo‘qotish funksiyasi:

$$L = -\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k} \log(\hat{y}_{i,k})).$$

Model baholash uchun quyidagi mezonlar ishlatiladi:

$$\text{Aniqlik (Accuracy)} = \frac{TP + TN}{TP + TN + FP + FN}, \qquad \text{Aniqlik (Precision)} = \frac{TP}{TP + FP},$$

$$\text{Qamrov (Recall)} = \frac{TP}{TP + FN}, \qquad F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

HISOBLASH NATIJALARI.

Ushbu tadqiqot natijalari matn hujjatlarini avtomatik tasniflashda taklif etilgan tizimning yuqori samaradorligini ko‘rsatdi. Eksperimentlar AG News to‘plami asosida o‘tkazilib, 120 ming hujjat o‘qitish va 7,600 hujjat sinov ma’lumotlari sifatida ishlatildi. Uch xil yondashuv sinovdan o‘tkazildi: an’anaviy mashina o‘rganishi usullari (Logistic Regression va Random Forest) TF-IDF xususiyatlari bilan, hamda chuqur neyron tarmoqlar (CNN + GRU) TF-IDF bilan birlashtirilgan holda.

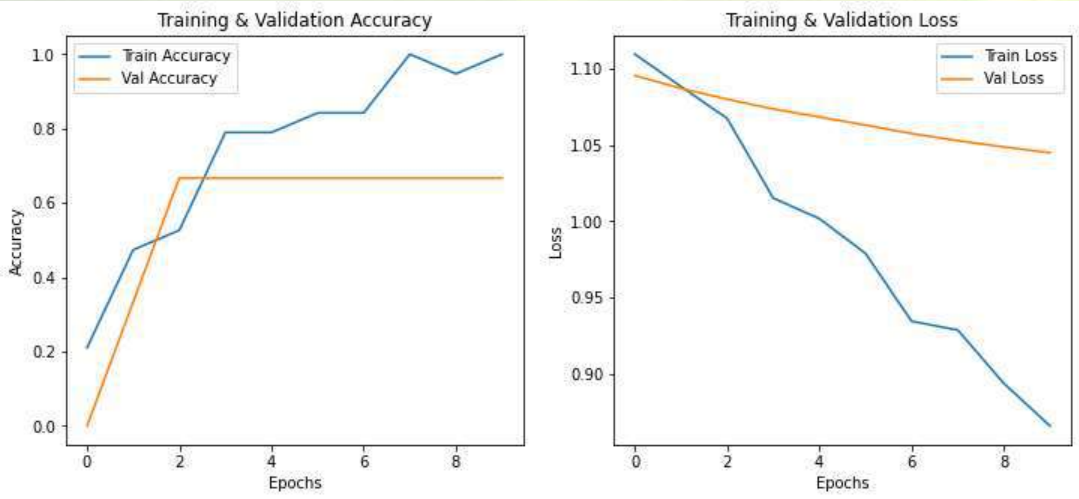
TF-IDF + Logistic Regression modeli hujjatlarni toifalashda 89,57% aniqlik ko‘rsatdi. Shu bilan birga, Random Forest modeli 90,12% aniqlikka erishdi (1-jadval). Ushbu natijalar shuni ko‘rsatadiki, an’anaviy usullar hujjatlarni toifalashda yetarlicha samarali bo‘lsa-da, ular murakkab va uzun matnlarda chuqur kontekstni o‘rganishda chegaralangan. Shuningdek, xarakteristik xususiyatlar (so‘zlar va ularning chastotalari) orqali ishlash modeli barqaror natija beradi, lekin semantik ma’lumotlarni chuqur anglash imkoniyati cheklangan (1-rasm).

1-jadval

Uch xil yondashuv sinovdan o‘tkazish natijalari

Model	Aniqlik (%)	Aniqlik (Precision)	Qamrov (Recall)	F1
TF-IDF + Logistic Regression	89.57	0.88	0.89	0.885
TF-IDF + Random Forest	90.12	0.89	0.90	0.895
CNN + GRU + TF-IDF	98.63	0.986	0.987	0.986

ISSN 3030-3907



1-rasm. Aniqlik va xatolik grafiklari

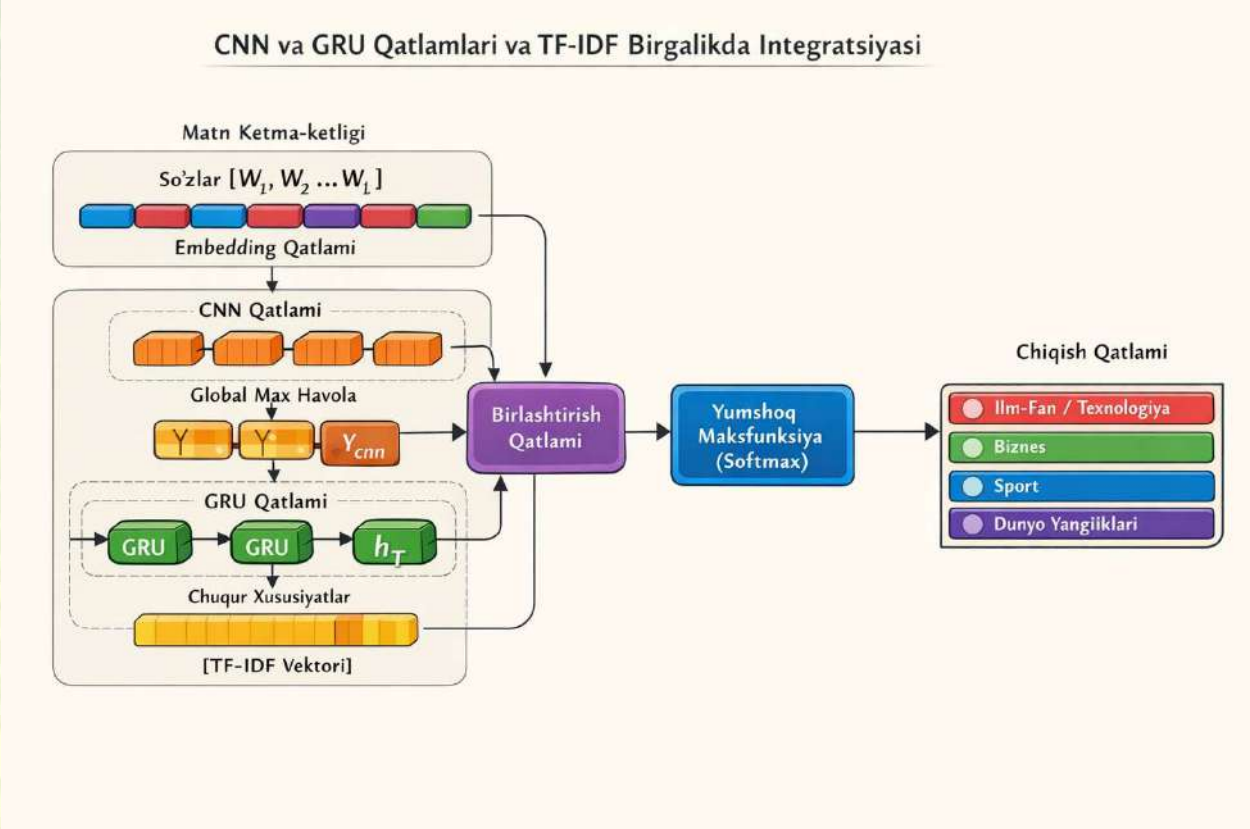
Ushbu tadqiqot natijalari matn hujjatlarini avtomatik tasniflashda taklif etilgan tizimning yuqori samaradorligini ko‘rsatdi. Eksperimentlar AG News to‘plami asosida o‘tkazilib, 120 ming hujjat o‘qitish va 7,600 hujjat sinov ma’lumotlari sifatida ishlatildi. Uch xil yondashuv sinovdan o‘tkazildi: an’anaviy mashina o‘rganishi usullari (Logistic Regression va Random Forest) TF-IDF xususiyatlari bilan, hamda chuqur neyron tarmoqlar (CNN + GRU) TF-IDF bilan birlashtirilgan holda.

TF-IDF + Logistic Regression modeli hujjatlarni toifalashda 89,57% aniqlik ko‘rsatdi. Shu bilan birga, Random Forest modeli 90,12% aniqlikka erishdi. Ushbu natijalar shuni ko‘rsatadiki, an’anaviy usullar hujjatlarni toifalashda yetarlicha samarali bo‘lsa-da, ular murakkab va uzun matnlarda chuqur kontekstni o‘rganishda chegaralangan. Shuningdek, xarakteristik xususiyatlar (so‘zlar va ularning chastotalari) orqali ishlash modeli barqaror natija beradi, lekin semantik ma’lumotlarni chuqur anglash imkoniyati cheklangan. CNN va GRU qatlamlarini TF-IDF bilan birlashtirish orqali ishlab chiqilgan model 98,63% aniqlikka erishdi. Bu shuni ko‘rsatadiki, chuqur neyron tarmoqlar hujjatlarni lokal va global kontekstlarda o‘rganish imkoniyatini beradi, TF-IDF esa matnning statistik xususiyatlarini qo‘shimcha ravishda hisobga oladi. Natijalar shuni ko‘rsatadiki, yondashuv semantik ma’lumot va so‘z chastotasi kombinatsiyasidan samarali foydalanadi (2-rasm).

Chalkashlik matritsasi tahlili ko‘rsatdiki, eng ko‘p xatoliklar “sport” va “biznes” toifalari o‘rtasida yuz beradi, bu mavzular o‘zaro leksik jihatdan yaqin bo‘lgani bilan izohlanadi. Shu bilan birga, “ilm-fan/texnologiya” va “dunyo yangiliklari” toifalari ancha barqaror aniqlanadi, bu ularning o‘ziga xos leksik va semantik xususiyatlariga bog‘liq. Ushbu tadqiqot natijalari avvalgi ishlarga nisbatan sezilarli yaxshilanishni ko‘rsatadi. Masalan, Heat et al. 94,9% aniqlikka erishgan bo‘lsa, Janani et al. 96,1% natija qayd etgan. Taklif etilgan CNN + GRU

ISSN 3030-3907

+ TF-IDF integratsiya yondashuvi esa 98,63% aniqlik bilan ularni ortda qoldirdi. Bu shuni ko‘rsatadiki, chuqur va sirtqi qatlamlarning integratsiyasi, hamda so‘z chastotasi asosidagi xususiyatlar birgalikda qo‘llanganda matn tasniflash samaradorligi sezilarli darajada oshadi.



2-rasm. CNN va GRU qatlamlarini TF-IDF bilan birlashtirish arxitekturasi

Shuningdek, trening va tekshiruv jarayonlari natijalaridan ko‘rinib turibdiki, tizim ortiqcha moslash (overfitting) holatiga duch kelmagan. Bu esa modelning turli xil hujjatlarda barqaror ishlashini ko‘rsatadi.Diskussiyada keltirilgan natijalar shuni ko‘rsatadiki, taklif etilgan tizim nafaqat yuqori aniqlik bilan ishlaydi, balki turli turdagi matn hujjatlarini avtomatik tasniflashga moslashuvchanlikka ega. Shu bilan birga, tizim real hayotdagi yangilik saytlarida, ijtimoiy tarmoqlarda, ilmiy va tijorat hisobotlarida hujjatlarni tez va ishonchli tarzda kategoriyalash imkonini beradi. Natijalar shuni ham ko‘rsatadiki, chuqur neyron tarmoqlarni TF-IDF bilan birlashtirish yondashuvi matnlarni tasniflashda an’anaviy yondashuvlarga nisbatan sezilarli ustunlik beradi. Bu metodologiya kelajakda boshqa tillar va sohalarda ham keng qo‘llanilishi mumkin.

XULOSA.

Ushbu tadqiqotda matn hujjatlarini avtomatik tasniflashga mo‘ljallangan avtomatik matn tasniflash tizimi ishlab chiqildi. Tadqiqotning asosiy maqsadi turli sohalardagi matnlarni yuqori aniqlik bilan tasniflash imkonini beruvchi tizim yaratish edi. Shu maqsadda Learning-based Text Classification (LbTC) algoritmi

ISSN 3030-3907

ishlab chiqildi, unda matnlarni oldindan tozalash, xususiyatlarni chiqarish va chuqur o‘rganish qatlamlari integratsiya qilindi. Tadqiqot davomida matnlar oldindan tozalandi, HTML va ortiqcha bo‘shliqlar olib tashlandi, maxsus belgilar chiqarildi, so‘zlar kichik harflarga aylantirildi va to‘xtovchi so‘zlar va lemmatizatsiya qilindi. So‘ngra matn xususiyatlarini chiqarish uchun TF-IDF va so‘z o‘rnatishlar (embedding) ishlatildi. Matn ketma-ketligi Konvolyutsion neyron tarmog‘i (CNN) va Eslab qoluvchi tarmoq (GRU) qatlamlaridan o‘tkazilib, chuqur xususiyatlar olinadi. Shu bilan birga, TF-IDF xususiyatlari Dense qatlam orqali ishlanadi va ikki yo‘l birlashtirilib, yakuniy softmax chiqish qatlamiga uzatiladi. Eksperimental natijalar shuni ko‘rsatdiki, taklif qilingan arxitektura an’anaviy mashina o‘rganishi modellari va alohida neyron tarmoqlar bilan solishtirganda eng yuqori aniqlikni berdi. Hujjatlarni tasniflashdagi umumiy aniqlik 98.63% ni tashkil qildi. Test natijalari shuningdek, modelning turli kategoriyalar bo‘yicha barqaror ishlashini va xatoliklarning minimal ekanligini ko‘rsatdi. Taklif qilingan tizimning ilmiy yangiligi shundaki, CNN va GRU qatlamlari TF-IDF xususiyatlari bilan integratsiya qilinib, matnlarni chuqur o‘rganish asosida toifalash imkonini berdi. Shu bilan birga, tizim katta hajmdagi matnli ma’lumotlar bilan samarali ishlashga qodir. Amaliy ahamiyati shundaki, bu tizim yangiliklar saytlarida, ilmiy maqolalar bazalarida, tijorat hisobotlarida va boshqa matnli resurslarda avtomatik tasniflashni tez va ishonchli amalga oshirish imkonini beradi. Kelajakda tizimni yanada rivojlantirib, oldindan o‘rganilgan modellarni integratsiya qilish va transfer o‘rganish metodlarini qo‘llash orqali aniqlik va samaradorlikni oshirish mumkin. Xulosa qilib aytganda, tadqiqot natijalari matn tasniflash bo‘yicha ilg‘or va samarali yondashuvni taqdim etadi hamda real hayotdagi turli ma’lumotlar bazalarida keng qo‘llanishi mumkinligini ko‘rsatadi.

ADABIYOTLAR

1. Huang, L., Ma, J., Chen, C., “Topic detection from microblogs using t-LDA and perplexity,” in *24th Asia-Pacific Software Engineering Conference Workshops (APSECW)*, IEEE, pp. 71–77, 2017.

2. Fang, W., Love, P. E., Luo, H., Ding, L., “Computer vision for behaviour-based safety in construction: A review and future directions,” *Advanced Engineering Informatics*, vol. 43, p. 100980, 2020.

3. Devlin, J., Chang, M. W., Lee, K., Toutanova, K., “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of NAACL 2019*, Minneapolis, MN, pp. 4171–4186, 2019.

4. Chen, N., Blostein, D., “A survey of document image classification: problem statement, classifier architecture and performance evaluation,” *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 10, pp. 1–16, 2007.

ISSN 3030-3907

5. Kumar, V., Recupero, D. R., Riboni, D., Helaoui, R., "Ensembling classical machine learning and deep learning approaches for morbidity identification from clinical notes," *IEEE Access*, vol. 9, pp. 7107–7126, 2020.
6. Zhong, B., Pan, X., Love, P. E., Ding, L., Fang, W., "Deep learning and network analysis: Classifying and visualizing accident narratives in construction," *Automation in Construction*, vol. 113, p. 103089, 2020.
7. Bertke, S. J., Meyers, A., Wurzelbacher, S., Bell, J., Lampl, M., Robins, D., "Development and evaluation of a naïve Bayesian model for coding causation of workers' compensation claims," *Journal of Safety Research*, vol. 43, pp. 327–332, 2012.
8. Mikolov, T., Chen, K., Corrado, G., Dean, J., "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.
9. Huang, J., Serb, A., Stathopoulos, S., Prodromakis, T., "Text classification in memristor-based spiking neural networks," arXiv:2207.13729, 2022.
10. Kumar, A., Bhalla, A., Garg, A., Modi, A., Shah, R. R., "Efficient text preprocessing for classification by learning from self-training," *IEEE Transactions on Knowledge and Data Engineering*, 2022.
11. Altınel, B., Ganiz, M. C., "Semantic text classification: A survey of past and recent advances," *Information Processing & Management*, vol. 54, pp. 1129–1153, 2018.
12. Fang, W., Luo, H., Xu, S., Love, P. E., Lu, Z., Ye, C., "Automated text classification of near-misses from safety reports: An improved deep learning approach," *Advanced Engineering Informatics*, vol. 44, p. 101060, 2020.
13. Janani, R., Vijayarani, S., "Automatic text classification using machine learning and optimization algorithms," *Soft Computing*, vol. 25, pp. 1129–1145, 2021.
14. Onan, A., "Hierarchical graph-based text classification framework with contextual node embedding and BERT-based dynamic fusion," *Journal of King Saud University— Computer and Information Sciences*, vol. 35, p. 101610, 2023.
15. Soni, S., Chouhan, S. S., Rathore, S. S., "TextConvONet: A convolutional neural network-based architecture for text classification," *Applied Intelligence*, vol. 53, pp. 14249–14268, 2023.

ISSN 3030-3907