



## TANQIDIY NAZAR, TAHLILIIY TAFAKKUR VA INNOVATSION G'UYALAR



### MASHINAVIY O'QITISH YORDAMIDA HUJJATLARNI AVTOMATIK TASNIFLASH USULLARI

**Jo'rayeva Asolat Akbar qizi.**

*Shahrisabz Davlat Pedagogika Instituti, Pedagogika Fakulteti Ta'lim va tarbiya nazariyasi va metodikasi yo'nalishi, Maktabgacha ta'lim Magistranti*

*E-mail: [asolat321123@gmail.com](mailto:asolat321123@gmail.com).*

**Fan o'qituvchisi; Qodirov Farrux Ergash o'g'li**

*Shahrisabz davlat pedagogika instituti "Matematika va ta'limda axborot texnologiyasi" kafedrasini mudiri i.f.d DSc, dotsent.*

*<https://orcid.org/0000-0002-4574-7728>, E-pochta: [fgodirov@shdpi.uz](mailto:fgodirov@shdpi.uz)*

**Annotatsiya** Ushbu maqolada mashinaviy o'qitish texnologiyalari yordamida hujjatlarni avtomatik tasniflash usullari tahlil qilinadi. Shuningdek, klassifikatsiya algoritmlari, semantik tahlil va tabiiy tilni qayta ishlash (NLP) texnologiyalarining samaradorligi ko'rib chiqiladi.

**Kalit so'zlar:** mashinaviy o'qitish, avtomatik tasniflash, hujjatlarni tahlil qilish, NLP, klassifikatsiya algoritmlari, matn qayta ishlash, sun'iy intellekt.

### МЕТОДЫ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ ДОКУМЕНТОВ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

**Жураева Асолат Акбар кизи.**

*Шахрисабзский государственный педагогический институт, Педагогический факультет Направление «Теория и методика обучения и воспитания»,*

*Магистрант дошкольного образования*

*F-mail: [asolat321123@gmail.com](mailto:asolat321123@gmail.com).*

**Преподаватель предмета; Кодиров Фаррух Эрғаш угли**

*Заведующий кафедрой «Математика и информационные технологии в образовании» Шахрисабзского государственного педагогического института,*

*доктор наук (DSc), доцент. ORCID: <https://orcid.org/0000-0002-4574-7728>*

*Электронная почта: [fgodirov@shdpi.uz](mailto:fgodirov@shdpi.uz)*

**Аннотация:** В данной статье анализируются методы автоматической классификации документов с использованием технологий машинного обучения. Также рассматривается эффективность классификационных алгоритмов, семантического анализа и технологий обработки естественного языка (NLP).

**Ключевые слова:** машинное обучение, автоматическая классификация, анализ документов, NLP, классификационные алгоритмы, обработка текста, искусственный интеллект.



# TANQIDIY NAZAR, TAHLILY TAFAKKUR VA INNOVATSION G'UYALAR



## METHODS OF AUTOMATIC DOCUMENT CLASSIFICATION USING MACHINE LEARNING

**Jo'rayeva Asolat Akbar qizi.**

*Shahrisabz State Pedagogical Institute, Faculty of Pedagogy*

*Specialty: Theory and Methodology of Education and Upbringing, Master's Student in  
Preschool Education E-mail: [asolat321123@gmail.com](mailto:asolat321123@gmail.com).*

**Subject teacher; Qodirov Farrux Ergash o'g'li**

*Head of the Department of "Mathematics and Information Technologies in Education"  
at Shahrisabz State Pedagogical Institute, Doctor of Science (DSc), Associate Professor.*

*ORCID: <https://orcid.org/0000-0002-4574-7728>*

*Email: [fgodirov@shdpi.uz](mailto:fgodirov@shdpi.uz)*

**Abstract:** *This article analyzes methods for automatic document classification using machine learning technologies. It also examines the effectiveness of classification algorithms, semantic analysis, and natural language processing (NLP) technologies.*

**Keywords:** *machine learning, automatic classification, document analysis, NLP, classification algorithms, text processing, artificial intelligence.*

Zamonaviy axborot texnologiyalari va raqamli hujjatlar sonining ortishi bilan hujjatlarni tezkor va samarali boshqarish muhim masalaga aylandi. An'anaviy qo'lda tasniflash jarayoni vaqt va resurs talab qiluvchi bo'lsa, mashinaviy o'qitish (Machine Learning) algoritmlari yordamida hujjatlar avtomatik tarzda kategoriyalarga ajratiladi [1, 20]. Ushbu maqolaning maqsadi — mashinaviy o'qitish usullari yordamida hujjatlarni tasniflash texnologiyalarini tahlil qilish, ularning samaradorligi va zamonaviy axborot tizimlaridagi qo'llanilishini ko'rsatishdir. Hujjatlarni avtomatik tasniflash uchun ko'plab mashinaviy o'qitish algoritmlari qo'llaniladi. Eng keng tarqalgani: Naive Bayes: matnlar ehtimollik asosida kategoriyalarga ajratiladi [2, 45]. Support Vector Machines (SVM): yuqori o'lchamli matn xususiyatlarini hisobga olib, hujjatlarni aniq klassifikatsiya qiladi [3, 50]. Decision Trees va Random Forest: matn xususiyatlariga asoslanib, hujjatlarni daraxt shaklida guruhlaydi [4, 60]. Tabiiy tilni qayta ishlash (NLP) yordamida matnli hujjatlar tahlil qilinadi. Bu jarayonda tokenizatsiya, stemming, lemmatizatsiya va TF-IDF usullari qo'llaniladi [5, 70]. Semantik tahlil orqali hujjatlarning mazmuniy o'xshashligi aniqlanadi. Bu esa shunchaki kalit so'zlarga emas, balki matnning umumiy mazmuniga asoslangan tasniflash imkonini beradi [6, 85]. Supervised learning (nazoratli o'qitish) usullari mashinaviy o'qitishda asosiy yo'nalish hisoblanadi. Hujjatlar oldindan belgilangan kategoriyalarga ajratilib, modelni o'rgatishda ishlatiladi. Shu orqali yangi hujjatlar avtomatik tarzda tegishli kategoriyalarga kiritiladi [7, 90]. Unsupervised learning (nazoratsiz o'qitish) esa klasterlash asosida ishlaydi. Hujjatlar mavzular bo'yicha guruhlanadi va ular orasidagi o'xshashlik asosida avtomatik tasniflanadi [8, 100].



## TANQIDIY NAZAR, TAHLILY TAFAKKUR VA INNOVATSION G'UYALAR



Deep Learning (chuqur o'rganish) texnologiyalari, xususan RNN va Transformer modellari, hujjatlarni yanada aniqlik bilan tasniflash imkonini beradi. Bu yondashuv katta hajmdagi matnlar bilan ishlashda samarali [9, 110].1. Matnlarni oldindan qayta ishlash (Preprocessing)

Hujjatlarni avtomatik tasniflashda matnlarni oldindan qayta ishlash muhim bosqich hisoblanadi. Tokenizatsiya, stop-so'zlarni olib tashlash, stemming va lemmatizatsiya kabi jarayonlar hujjatlarning asosiy mazmunini saqlab, tasniflash algoritmlari samaradorligini oshiradi [2, 150].

Hujjatlarni tasniflash jarayonida matndan olinadigan xususiyatlar (features) sonini kamaytirish hisoblash resurslarini tejaydi va modelning aniqligini oshiradi. Buning uchun TF-IDF, word embeddings (Word2Vec, GloVe) va BERT kabi metodlar ishlatiladi [5, 155].

Hujjatlarning mazmuni va kontekstini tushunish uchun semantik tahlil va tabiiy tilni qayta ishlash (NLP) texnologiyalari qo'llaniladi. Bu nafaqat kalit so'zlarga, balki matnning umumiy mazmuniga asoslangan tasniflash imkonini beradi [6, 160]. Ba'zi hollarda unsupervised learning (nazoratsiz o'qitish) yordamida hujjatlar o'xshash mavzular bo'yicha guruhlanadi. Keyin bu klasterlar asosida tasniflash algoritmlari yanada aniq ishlaydi. Bu yondashuv katta hajmdagi Chuqur o'rganish (Deep Learning) texnologiyalari, xususan RNN, LSTM va Transformer modellari, hujjatlarni yuqori aniqlik bilan tasniflash imkonini beradi. Ushbu algoritmlar matnning kontekstini hisobga olib, murakkab tuzilishga ega hujjatlarni ham aniq tasniflaydi [9, 170].

### Modelning baholanishi

Hujjatlarni avtomatik tasniflashda samaradorlikni baholash muhim. Accuracy, Precision, Recall, F1-score kabi ko'rsatkichlar algoritmning ishlash sifatini tahlil qilish imkonini beradi. Shu bilan birga, Cross-Validation usuli orqali modelning barqarorligi sinovdan o'tkaziladi [2, 155]. Amaliy qo'llanish sohalari Hujjatlarni avtomatik tasniflash tizimlari korporativ hujjatlarni boshqarish, ilmiy maqolalar arxivini tartiblash, moliyaviy va hukumat hujjatlarini boshqarish hamda kutubxonalar va onlayn resurslar bilan ishlashda qo'llaniladi [3, 175].

Bulutli va onlayn tizimlar Bulutli platformalarda hujjatlarni avtomatik tasniflash tizimlari katta hajmdagi ma'lumotlarni tezkor boshqarish va ko'p foydalanuvchi rejimida ishlash imkonini beradi. Bu esa tizimning skalabilitesini oshiradi [6, 180].

Gibrid yondashuvlar (Hybrid Approaches) Hujjatlarni tasniflash jarayonini yanada mukammal qilish uchun NLP + Deep Learning + Clustering kombinatsiyasi ishlatiladi. Bu yondashuv nafaqat aniqlikni oshiradi, balki tizimning adaptivligini ham kuchaytiradi [5, 185].

Real vaqt rejimida tasniflash Ba'zi tizimlar yangi kelayotgan hujjatlarni real vaqt rejimida tasniflaydi. Bu, ayniqsa, yangiliklar, elektron pochta va moliyaviy hisobotlar bilan ishlashda samarali [7, 190]. Foydalanuvchi tajribasini yaxshilash Avtomatik tasniflash tizimlari foydalanuvchiga kerakli hujjatlarni tez va aniq topish imkonini beradi, bu esa ish jarayonini tezlashtiradi va resurslardan optimal foydalanishga yordam beradi [8, 195].12. Hujjatlarning turli formatlarini qo'llab-quvvatlash Mashinaviy o'qitish tizimlari faqat matnli hujjatlar bilan cheklanmaydi. PDF, Word, HTML, XML va boshqa formatdagi hujjatlarni ham tahlil qilib,



tasniflash imkoniyatiga ega. Buning uchun formatga mos parser va ekstraktorlar ishlatiladi [2, 200]. Ko‘p tilli hujjatlar bilan ishlash Global tashkilotlarda hujjatlar turli tillarda bo‘lishi mumkin. NLP va tarjima algoritmlari yordamida tizim turli tillardagi matnlarni bir xil kategoriyaga to‘g‘ri tasniflay oladi [3, 205].

Avtomatik annotatsiya va teglar yaratish Mashinaviy o‘qitish tizimlari hujjatlarni avtomatik tarzda teglar bilan belgilashi mumkin. Bu esa qidiruv tizimlarining samaradorligini oshiradi va foydalanuvchiga kerakli hujjatlarni tez topish imkonini beradi [5, 210]. Ma’lumotlar xavfsizligi va maxfiylik Hujjatlarni avtomatik tasniflash tizimlari korporativ ma’lumotlarni himoya qilish va maxfiylikni ta’minlashga ham xizmat qiladi. Foydalanuvchilarga ruxsat berilgan hujjatlar va kategoriyalar orqali materiallarga kirish nazorat qilinadi [6, 215].

Foydalanish sohalari Korporativ sektor: xodimlar va bo‘limlar uchun hujjatlarni tezkor boshqarish. Kutubxonalar va ilmiy arxivlar: maqolalar, kitoblar va dissertatsiyalarni guruhlash. Hukumat va moliya idoralari: elektron hujjatlarni tartiblash, monitoring va audit jarayonlarini optimallashtirish [3, 220].

Adaptiv va real vaqt rejimidagi tizimlar Yangi kelayotgan hujjatlar tizim tomonidan avtomatik tarzda tasniflanadi va foydalanuvchiga darhol taqdim etiladi. Bu, ayniqsa, elektron pochta, yangiliklar, moliyaviy hisobotlar va onlayn arxivlarda muhim ahamiyatga ega [7, 225]. Mashinaviy o‘qitish modellari optimizatsiyasi Hyperparameter tuning yordamida model samaradorligi oshiriladi. Ensemble learning orqali bir nechta algoritmlarning natijalari birlashtiriladi va aniqlik oshadi. Transfer learning yordamida oldindan tayyorlangan modellardan foydalangan holda yangi hujjatlar tezkor tasniflanadi [9, 230]. Foydalanuvchi interfeysi va vizualizatsiya Hujjatlar tasnifi natijalari vizual diagramma, graf va mavzu bulutlari orqali ko‘rsatiladi. Bu foydalanuvchi uchun intuitiv va tushunarli bo‘lib, qidiruv jarayonini tezlashtiradi [8, 235].

Kelajak istiqbollari Mashinaviy o‘qitish va NLP algoritmlari yanada optimallashtiriladi. Avtomatik tasniflash tizimlari bulutli va mobil platformalarda kengaytiriladi. Hujjatlar va resurslar bilan ishlash real vaqt rejimida yanada tezkor va samarali bo‘ladi. Evaluatsiya va samaradorlik o‘lchovlari sifatida Accuracy, Precision, Recall va F1-score keng qo‘llaniladi. Bu ko‘rsatkichlar algoritmlarning aniqligini baholash imkonini beradi [2, 95]. Hujjatlarni avtomatik tasniflash tizimlari korporativ axborot tizimlarida, kutubxonalar va ilmiy arxivlarda, shuningdek hukumat va moliya idoralari keng qo‘llaniladi. Bu tizimlar katta hajmdagi hujjatlarni tezkor tarzda tartiblash, qidirish va boshqarishni ta’minlaydi [3, 120]. Hybrid yondashuvlar — masalan, NLP + Deep Learning + Clustering — hujjatlarni yanada aniq va samarali tasniflash imkonini beradi. Shu orqali tizim foydalanuvchiga kerakli hujjatlarni tez va oson topish imkonini yaratadi [5, 130].

Bulutli platformalarda hujjatlarni avtomatik tasniflash tizimlari resurslarni samarali boshqarish va ko‘p foydalanuvchi rejimida ishlash imkonini beradi [6, 140].

Xulosa qilib aytganda, mashinaviy o‘qitish yordamida hujjatlarni avtomatik tasniflash tizimlari zamonaviy axborot tizimlarining ajralmas qismiga aylanmoqda. Ular hujjatlarni tez,





## TANQIDIY NAZAR, TAHLILIIY TAFAKKUR VA INNOVATSION G'OYALAR



aniq va samarali tarzda kategoriyalarga ajratib, foydalanuvchiga kerakli ma'lumotni tez topish imkonini beradi. Xulosa qilib aytganda, mashinaviy o'qitish yordamida hujjatlarni avtomatik tasniflash zamonaviy axborot tizimlarining eng muhim tarkibiy qismlaridan biridir. Ushbu texnologiya yordamida hujjatlar katta hajmga ega bo'lsa ham, tezkor va aniq tarzda kategoriyalarga ajratiladi, foydalanuvchi esa kerakli ma'lumotni minimal vaqt ichida topa oladi.

Supervised va unsupervised learning algoritmlari, shuningdek, chuqur o'rganish (Deep Learning) texnologiyalari hujjatlarni yuqori aniqlik bilan tasniflash imkonini beradi. Shuningdek, NLP va semantik tahlil texnologiyalari matn mazmunini chuqur tushunib, tasniflash jarayonini optimallashtiradi. Klasterlash va hybrid yondashuvlar esa tizimning adaptivligi va samaradorligini oshiradi. Hujjatlarni avtomatik tasniflash tizimlari faqat korporativ va ilmiy sohalarda emas, balki hukumat, moliya, kutubxona va onlayn platformalarda ham keng qo'llaniladi. Bulutli platformalar, real vaqt rejimi va ko'p tilli qo'llab-quvvatlash tizimlari ushbu jarayonni yanada tezkor va qulay qiladi.

Kelajakda avtomatik tasniflash tizimlari yanada rivojlanib, foydalanuvchi tajribasini oshiradi, axborot resurslarini samarali boshqarishni ta'minlaydi va ta'lim, ilm-fan hamda ish jarayonlarida qo'llanilish imkoniyatlarini kengaytiradi. Shu bilan birga, mashinaviy o'qitish asosidagi hujjatlarni tasniflash tizimlarini tadqiq qilish va amaliyotga joriy etish zamonaviy axborot texnologiyalari va raqamli transformatsiya sohasining eng dolzarb yo'nalishlaridan biri hisoblanadi.

Supervised va unsupervised learning algoritmlari, shuningdek Deep Learning texnologiyalari hujjatlarni yuqori aniqlik bilan tasniflash imkonini beradi. NLP va semantik tahlil esa matn mazmunini chuqur tushunib, algoritmlarning samaradorligini oshiradi. Kelajakda bu texnologiyalar yanada rivojlanib, katta hajmdagi raqamli hujjatlar bilan ishlashni tezlashtiradi, axborot tizimlarining interaktivligini va foydalanuvchi tajribasini oshiradi. Shu sababli, mashinaviy o'qitish yordamida hujjatlarni avtomatik tasniflash tizimlarini o'rganish va amaliyotga joriy etish zamonaviy axborot texnologiyalarining eng dolzarb yo'nalishlaridan biri hisoblanadi.

### Foydalanilgan adabiyotlar

1. Qodirov, F. "" QR-KOD TEXNOLOGIYASI ASOSIDA ELEKTRON KUTUBXONA TIZIMINI DASTURIY VA APPARAT TAMINOTINI YARATISH". MUHAMMAD AL-XORAZMIY NOMIDAGI TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI QARSHI FILIALI." (2021).

2. Qodirov, F. E., O. D. Doniyorov, and H. Shokirov Sh. "Basic Concepts Of Information Security In Information Systems. Wide Threats And Their Consequences." КОНЦЕПЦИИ УСТОЙЧИВОГО РАЗВИТИЯ НАУКИ В СОВРЕМЕННЫХ УСЛОВИЯХ (2021): 153-155.

