

UO'K 004.9

MA'LUMOTLARNI MOSLASHTIRISHNING ZAMONAVIY YONDASHUVLARI

Nurmamatov M.Q.¹

¹ Sharof Rashidov nomidagi Samarqand davlat universiteti, Samarqand, O'zbekiston
mehriddinnur@gmail.com

Annotatsiya. Bugungi kunda axborot xajmining keskin suratlarda oshib borayotganligi hamda har bir shaxs uchun turli ma'lumotlar bazasidagi yozuvlardan foydalangan holda xulosalar berish jarayoni takomillashib bormoqda. Turli ma'lumotlar bazalaridagi yozuvlarni moslashtirish orqali yagona bazaga yig'ish orqali qarorlar qabul qilish ish jarayonlarni optimallashtirish hamda iqtisodiy samaradorlikka egadir. Tibbiyot, Online savdolar, Bandlik tizmlarida bu kabi moslashtirishlar vaqt va iqtisodiy ko'rsatkichlarni yaxshilaydi. Shu sababdan ushbu maqolada turli ma'lumotlar bazasidagi yozuvlarni moslashtirishning usul, model va yondashuvlari o'rganilgan.

Kalit so'zlar: K-means, mashinaviy o'qitish, bandlik, ma'lumotlar bazasi, bloklash, Q-gramm, ekspert baholash.

I. KIRISH

Dunyoning ko'plab rivojlangan davlatlarida aholini ro'yxatga olish har o'n yilda bir marta amalga oshiriladi. Jumladan yurtimizda ham yaqin 2 yil ichida aholini ro'yxatga olish rejalashtirilgan. Har bir davlat uchun aholini ro'yxatga olish katta statistik jarayonlardan biri hisoblanadi. Aholini ro'yxatga olish 2 usuldan biri orqali amalga oshirish mumkin:

1. An'anaviy usul – mahallabay har bir xonadon bo'yicha amalga oshiriladi.

2. Ma'muriy usul – aholini ro'yxatga olish ma'lumotlari davlat boshqaruvi ma'lumotlar bazalaridan bosqichma-bosqich olinadi. Bunda tashkiliy jarayonlar uchun harajatlar va qo'l mehnatini kamaytirish, aholini ro'yxatga olish ma'lumotlaridan kengroq foydalanish imkoniyatini ta'minlash kabi qulayliklari bilan ajralib turadi.

Respublikamizda aholini ro'yxatga olish an'anaviy usul, ya'ni mahallabay har bir xonadon bo'yicha so'rovnomasida amalga oshirilmog'na. Lekin ro'yojlangan davlatlarning askariyatida ma'muriy usuldan foydalanishni boshladilar bunda statistikani olishda yaxshi samaradorlikka erishish mumkin va an'anaviy usulda ma'lumotlarining yetishmasligining oldi olinadi. Misol uchun, O'zbekistonda statistika ma'lumotlariga ko'ra 1-may holatiga aholining 37 milliondan oshganligini ko'rish mumkin. Aholini ro'yxatga olish jarayonini ma'muriy usuldan foydalanish orqali amalga oshirish iqtisodiy samarador hisoblanadi. Yuqoridagilardan kelib chiqib aholini ro'yxatga olish uchun mas'ul bo'lgan O'zbekiston respublikasi prezidenti

huzuridagi statistika agentligi ro'yxatga olish hamda ma'lumotlarini yig'ish jarayonini gibrid usulga (an'anaviy va ma'muriy usulning kombinatsiyasi) aylantirish orqali amalga oshirsa ko'zlangan maqsadga vaqt jihatdan ham iqtisodiy jihatdan ham samarali yechim qilinishiga olib keladi [1,2].

II. ASOSIY QISM

Turli ma'lumotlar bazalaridagi ma'lumotlarni moslashtirish

Mavjud bo'lgan yozuvlarni moslashtirishni tezlashtirish uchun odatda qo'llaniladigan usullardan biri bu standart bloklashdir [3]. Standart bloklash-bu bloklash kaliti yordamida o'xshash yozuvlarni guruhlashdan iborat indekslash texnikasi hisoblanadi. Bizga ikkita ma'lumotlar bazalarida mavjud yozuvlar berilgan bo'lsin.

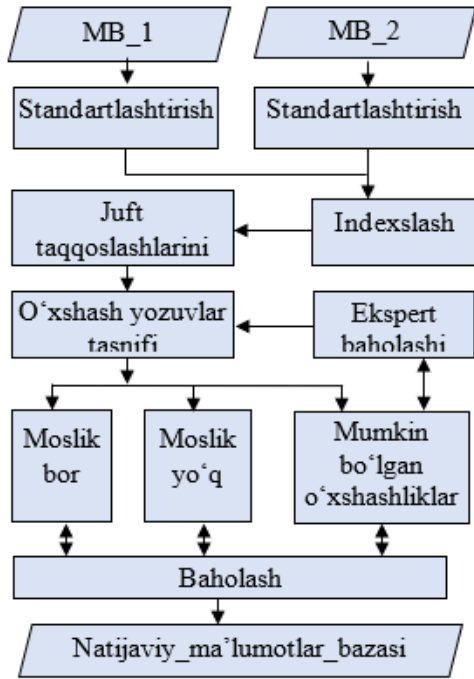
1-rasmda moslashtirish jarayonini ko'rish mumkin. Bunda moslashtirish bir xil kalitga ega yozuvlarni guruhlash orqali amalga oshiriladi. Bunday holatda millionlab yozuvlar mavjud bo'lgan jadvallarda mos yozuvlarni birlashtirish orqali qidiruv va qayta ishlash jarayoni bir muncha tezlashganini ko'rishimiz mumkin [7].

Ma'lumotlarning moslik yondashuvlaridan biri ma'lumotlar juftligini mos yoki mos kelmaydigan deb tasniflashdir. Buning uchun biz mashinani o'qitish modellarining bir nechta turlaridan foydalanishimiz mumkin. Bunga nazorat ostida o'qitish, yarim nazorat ostida o'qitish, faol o'qitish yoki nazoratsiz o'qitish kabi modellar orqali erishish mumkin.

Nazorat ostida o'qitishda ma'lumotlar bazasida aniqlangan tanlanma beriladi. Bunda ma'lumotlar bazasida yozuvlar mos va mos kelmaydigan

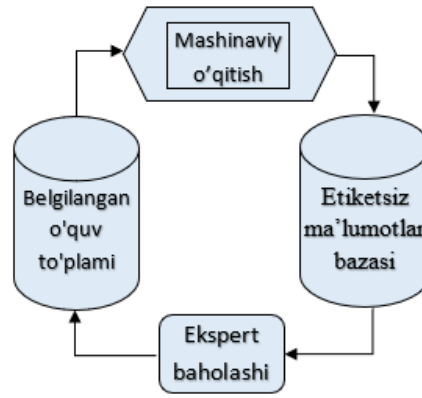
Ma'lumotlarni moslashtirishning zamonaviy yondashuvlari

juftliklardan iborat bo'ladi, natijada bu ma'lumotlar bazasi asosida qarorlar daraxtlari, logistik regressiya va boshqa chuqur o'qitishlar orqali natijalar shakllantiriladi. Yana bir jihati shundan iboratki, Yarim nazorat ostida o'qitishda o'quv ma'lumotlari nafaqat etiketlangan ma'lumotlarni, balki etiketlanmagan ma'lumotlarni ham o'z ichiga oladi.



1-rasm. Ikkita ma'lumotlar bazasidagi maydonlarni moslashtirish sxemasi

Faol o'qitish - bu etiketli misollarning boshlang'ich kichik o'qitish ma'lumotlari to'plamiga asoslangan o'rganish usuli. O'qitish algoritmi dastlab etiketli ma'lumotlar to'plami bilan o'qitiladi va keyin yorliqlar uchun foydalanuvchiga so'rovlarni so'raydi (2-rasm).



2-rasm. Faol o'qitishning siklik ko'rinishi

Nazoratsiz o'qitishda faqat etiketsiz ma'lumotlardan foydalanadi. Bunda ma'lumotlar asosida modellar yaratish natijasida ularni guruhlashga va sinflarni ifodalashi mumkin bo'lgan klasterlarni shakllantirishga asoslanadi. Bunga misol sifatida K-Meansni keltirish mumkin [9].

Logistik regressiya algoritmi odatda ikkita mumkin bo'lgan natija (faqat ikkita mumkin bo'lgan natijaga ega bo'lgan tasniflash) ishlatiladigan nazorat qilinadigan o'qitish usuli hisoblanadi [10]. Bunda tahlil qilinayotgan yozuvlar aniqlangan aniq sinfga mos kelishi ehtimolligi hisoblanadi va natijada 0 yoki 1 qiymat qaytariladi.

$$\alpha(t) = \frac{1}{1 + e^{-t}} \tag{1}$$

Yuqoridagi (1) logistik funksiyaning kirish qiymatlari t , og'irliklar yordamida chiziqli tarzda birlashtiriladi, odatda β sifatida ifodalanadi. Quyidagi 2-formulada bir xil logistik funktsiyani t kirishda β_1 va β_2 og'irliklari bilan ifodalaydi. Bunda og'irlik koeffitsientlarni o'quv tanlanma yordamida maksimal ehtimollikni baholash bilan hisoblash mumkin.

$$d(i, j) = \min \begin{cases} d(i-1, j-1) + c(x_i, y_j) & \text{nusxalash yoki almashtirish} \\ d(i-1, j) & x_i \text{ ni o'chirish} \\ d(i, j-1) & y_j \text{ ni almashtirish} \end{cases} \tag{2}$$

$$c(x_i, y_j) = 0.$$

Agar $x_i = y_j$, u holda 1, aks holda:

$$d(0, 0) = 0; d(i, 0) = i; d(0, j) = j.$$

Ikkita satr orasidagi masofani tahrirlash bilan hisoblangan o'xshashlik balli tasvirlangan (1-jadval). Bekhzod va Behzod o'rtasidagi

o'xshashlik bitta, shu sababdan satrni boshqasiga moslashtirish uchun faqat bitta operatsiya kerak. Quyidagi misolda Bekhzod yozuvida k ni o'chirish orqali amaliyot bajarilmoqda.

Q-grammalar bir satrning q belgilarining ketma-ketligi. Masalan, bigrammani (2 o'lchamdagi q-gramma) olish uchun, 2

o'Ichamdagi oyna yozuvi bo'ylab siljiydi natijada belgilar bigramlarni hosil qilinadi.

1-jadval. Yozuvlar orasidagi o'xshashlikni tahrirlash

		B	E	K	H	Z	O	D
	0	1	2	3	4	5	6	7
B	1	0	1	2	3	4	5	6
E	2	1	0	1	2	3	4	5
H	3	2	1	0	1	2	3	4
Z	4	3	2	1	0	1	2	3
O	5	4	3	2	1	0	1	2
D	6	5	4	3	2	1	0	1

Masalan, "muhandis" so'zining bigramlari [mu],[uh], [ha], [an], [nd], [di], [is].

$|s_1|$ va $|s_2|$ satrlar o'rtasidagi moslik koeffitsiyenti Jaccard formulasi (3-formula) yordamida topiladi:

$$Jaccard(s_1, s_2) = \frac{|s_1 \cap s_2|}{|s_1 \cup s_2|}, \quad (3)$$

bu yerda s_1 va s_2 uchun, q-grammlari umumiy sonini $|s_1 \cap s_2|$ hamda $|s_1 \cup s_2|$ esa q-grammlar birlashmasining hajmini bildiradi.

Yozuvlarni tahlil qilish birinchi navbatda ism va familiyalarni solishtirish orqali amalga oshiriladi. Buning uchun Jaro metrikasidan foydalaniladi:

$$Jaro(s_1, s_2) = \frac{1}{3} \left(\frac{c}{|s_1|} + \frac{c}{|s_2|} + \frac{c-t}{c} \right). \quad (4)$$

$$sim_{kum_abs}(k_1, k_2) = \begin{cases} 1.0 - \left(\frac{|k_1 - k_2|}{k_{max}} \right) & \text{agar } |k_1 - k_2| < k_{max} \\ 0.0 & \text{aksholda} \end{cases} \quad (6)$$

Agar o'rganish usulida etiketlanmagan ma'lumotlardan ko'proq noaniq bo'lganlarni tanlasa, u tasniflashni tezlikda yaxshilaydi va mustahkamlaydi. Yuqoridagi noaniq holatlarni aniqlash uchun bir-biridan farq qiladigan shu bilan birga shunga o'xshash aniqlikka ega tasniflashlar guruhi ishlatiladi. Tadqiqot davomida yozuvlarni mos keladigan, mos kelmaydigan va potensial mos keladiganlarga tasniflash uchun Threshold-Based-Classification usulini tanlab olindi [6]. Bunda ikki yozuv o'rtasida o'xshashlik metrikasi bilan oldin taqqoslangan maydonlarning barcha o'xshashlik ballarini yig'ish va agar bu yuqori chegaradan yuqori bo'lsa, u mos keladi, agar pastki chegaradan past bo'lsa, u mos kelmaydi degan xulosani berishdan iborat bo'ladi. Algoritm tomonidan aniqlangan ball chegaralar orasida

Jaro modifikatsiyasining yan bir Jaro-Winkler hisoblanib u- ism va familiyalarni solishtirishda kamroq xatoga yo'l qo'yishi bilan ajralib turadi. Bunda ikkita yangi PL va PW parametrlarni qo'shish orqali hosil qilinadi, bu yerda PL satrlar orasidagi eng uzun umumiy prefiksning uzunligi, PW prefiksning og'irligi hisoblanadi (3):

$$Jaro-Winkler(s_1, s_2) = (1 - PL \times PW) \times jaro(s_1, s_1) + PL \times PW. \quad (5)$$

Hozirgacha taqdim etilgan metrikalar xarakterga asoslangan. Soundex mag'lubiyatning fonetik ifodasiga asoslangan bo'lib, odatda nomlarni mos keltirish uchun ishlatiladi. Berilgan mag'lubiyat uchun metrik shunday ishlaydi.

2.1 O'xshashlikni aniqlash samaradorligi. Tadqiq qilinayotgan ma'lumotlar bazalarida mavjud yozuvlarning nusxalarini aniqlashda samaradorlik katta ahamiyatga ega. Yozuvlarni samarali bog'lash uchun moslashtirilgan saralangan qo'shnihilik usullari haqida [4] da batafsil ma'lumot berilgan.

Yozuvlarni moslashtirishdagi yana bir muhim bosqichlardan bir sana va vaqt bilan bog'liq ustunlarni solishtirishdan iborat hamda takrorlarni aniqlash hisoblanadi. Sana va vaqtlarni solishtirish [5] da batafsil keltirilgan bo'lib bunda (6) formulada k_{max} o'zgaruvchisi k_1 va k_2 kunlar o'rtasidagi maksimal farqni ya'ni chegarani ifodalaydi:

bo'lsa, u potensial mos keladi hamda natijani ma'muriyat yoki mutaxassis tomonidan tekshirishga yuboriladi.

2.2 O'xshashliklarni aniqlash yondashuvlari. Ma'lumotlar bazasidagi yozuvlarning mosligini tekshirishning oddiy va samarali algoritmlardan biri hisoblangan Logistik regressiya tahlil qilindi. Logistik regressiya algoritmlarining tanlanishining aosiy sablaridan biri tez ishlashi bilan bir qatorda dastlabki o'qitishlardan keyin ijobiy natijalarni ko'rsatishi bilan ajralib turadi. Ma'lumotlar bazasidagi millionlab yozuvlarni solishtirib chiqishda taqqoslashlar soni oshib ketishi bilan murakkablik ham ortib boradi [8].

Izlanayotgan yechim ikki bosqichda ega:

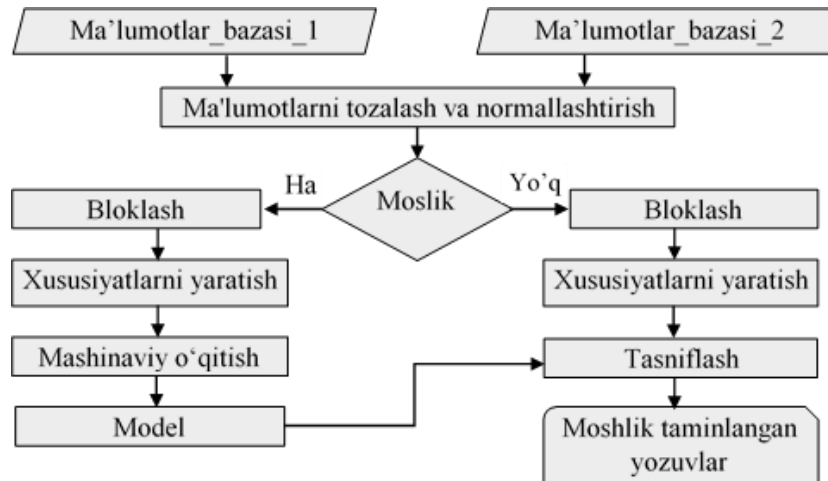
Ma'lumotlarni moslashtirishning zamonaviy yondashuvlari

- o'rganish yoki o'qitish bosqichi, oldindan aniqlangan etiketli ma'lumotlarni logistik regressiya orqali o'qitilib maslarning modeli shakllantiriladi;

- sinov yoki tasniflash bosqichi, bunda ma'lumotlar mos yoki moslik yo'q guruhlariga ajratiladi.

Tadqiq etilayotgan usulning arxitekturasi 3-rasmda keltirilgan. Jarayon ma'lumotlar bazalari

juftlarini tanlashdan boshlanadi, masalan, turli ikkita ma'lumotlar bazasidagi yozuvlardan iborat jadval berilgan bo'lsin (2-jadval). Qayta ishlash maqsadida bir xil maydonlar ajratib olinadi. Bunda ma'lumotlarni tozalash va normallashtirish amalga oshirishimiz kerak. Chunki biz faqat umumiy maydonlarni tanlashimiz kerak va bu maydonlar bir xil ko'rinishlarga ega bo'lishi kerak.



3-rasm. Yozuvlarni moslashtirish tizimining arxitekturasi

III. NATIJALAR TAHLILI

2-jadvalning yozuvlarini mos yoki mos bo'lmagan juftliklarini aniqlash imkonini beradigan matematik modelni qurib olishimiz kerak bo'ladi. Shu sababdan modelni qurish uchun etiketli ma'lumotlardan foydalaniladi.

Etiketli ma'lumotlar umumiy kalit orqali ikkita ma'lumotlar bazalarini qo'shish orqali olinadi. Quyidagi jadvalda umumiy kalit sifatida fuqoroning ID raqamidan foydalanilgan holda jadvallarni birlashtirildi. Natijada bizda ikkita yozuvlar maydoni hosil bo'ladi modelni o'qitish uchun mos kelishlar ($a1$ va $b1$, $a2$ va $b2$ juftliklar) va mos kelmagan yozuvlar ($a3$ va $b3$, $a4$ va $b4$).

2-jadval. Ma'lumotlar_bazasi_1 va ma'lumotlar_bazasi_2 dagi yozuvlar

Yozuvlar	ID	Ismi	Familiyasi	Tug'ulgan_yili
$a1$	1896090023	Ali	Abdullayev	10.11.1989
$a2$	1912090023	Ali	Salimov	10.11.1989
$a3$	1966060023	Behzod	Avazov	31.10.1996
$a4$	1996096623	Qodir	Murodov	12.01.1999

Yozuvlar	ID	Ismi	Familiyasi	Tug'ulgan_yili	
$b1$	1896090023	31011896090023	Ali	Abdullayev	10.11.1989
$b2$	1912090023	31011896090023	Ali	Salimov	10.11.1989
$b3$	1966060023	33110966060023	Behzod	Qodirov	31.10.1996
$b4$	1996096623	31201996096623	Jasur	Murodov	12.01.1999

O'qitish jarayoni boshlangandan keyin ikkita ma'lumotlar bazasi o'rtasida juft yozuvlarni yaratish uchun standart blokirovkalash usulini qo'llaymiz. Bu orqali mosliklarni saqlab qolamiz va bitta blok ichida ko'pi bilan bitta haqiqiy moslikka ega yoki 0 ta va undan ortiq mos

kelmasliklar bo'lishi mumkin. Tadqiq qilinayotgan blokirovka mezonlari Ismi+Tug'ilgan_yil yozuvlaridan foydalaniladi. 3-jadvalda $r1$ dan $a4$ gacha va $b1$ dan $b4$ gacha bo'lgan yozuvlar 2-jadvalda yozuvlar asosida hosil qilindi. Ushbu misolda, ID kaliti orqali mos keladigan barcha

yozuvlar qanday yaratilganligini ko'rsatish uchun bir xil blokirovka kalitiga ega yoki ega emas ekanligi aniqlanadi. Shunday qilib, $a1$ va $a2$

yozuvlari $b1$ va $b2$ bilan bir xil blokga o'tadi. Juft yozuvlar bu maydonlar orasidagi o'xshashlik qiymatlari sifatida ifodalanadi.

3-jadval. Bloklash kaliti yordmida ma'lumotlar_bazasi_1 va ma'lumotlar_bazasi_2 dagi yozuvlar

Yozuvlar	Bloklash kaliti
$a1$	Ali10111989
$a2$	Ali10111989
$a3$	Behzod31101996
$a4$	Jasur12011999

Yozuvlar	Bloklash kaliti
$b1$	Ali10111989
$b2$	Ali10111989
$b3$	Behzod31101996
$b4$	Jasur12011999

Tadqiqotda foydalanadigan o'xshashlik ko'rsatkichi masofani tahrirlash hisoblanadi. U 4-jadvalda bir xil blokda bo'lgan yozuvlar juftligi

orasidagi masofani tahrirlash bo'yicha ballar asosida $r1$, $s2$ va $r2$, $s1$ yozuvlarning mosliklar bahosi keltirilgan.

4-jadval. Taqqoslangan yozuv maydonlarining o'xshashlik ko'rsatkichlari

Yozuvlarni taqqoslash	Ismi	Familiyasi	Tug'ulgan_yili	Holati
$(a1, b1)$	0	0	0	Moslik bor
$(a1, b2)$	0	3	0	Moslik yo'q
$(a2, b1)$	0	3	0	Moslik yo'q
$(a2, b2)$	0	1	0	Moslik bor

Tasniflashni o'qitishning birinchi bosqichning oxirgi qadami (3-rasmdagi O'qitish qadami) bu mos kelmaydigan yozuvlarni klassifikatsiya qila oladigan modelni hosil qilish hisoblanadi. Klassifikatsiya bosqichi mos kelmaydigan yozuvlarga bir xil bloklash usulini, ya'ni Standart Bloklashni va bir xil bloklash mezonlarini qo'llash

bilan boshlanadi. Mos kelmaydigan yozuvlar - bu umumiy kalit orqali mos kelmagan har bir ma'lumotlar bazasidan qolgan yozuvlardir. Bloklash orqali solishtirish uchun yozuvlar juftlashtiriladi va o'qitish bosqichidagi kabi har bir maydon juftligi uchun tahrirlash masofasi qo'llaniladi, 5-jadval.

5-jadval. Taqqoslanmagan yozuv maydonlarining o'xshashlik ko'rsatkichlari

Yozuvlarni taqqoslash	Ismi	Familiyasi	Tug'ulgan_yili	Holati
$(a3, b3)$	0	1	0	?
$(a4, b4)$	0	3	0	?

Yuqoridagilardan kelib chiqib model satr o'xshashlik ko'rsatkichi bilan qaytarilgan ushbu qiymatlarni qabul qilishi va har bir yozuv juftligini mos keladigan yoki mos kelmaydigan (Tasniflash bosqichi) tasniflashi mumkin, 6-jadval.

6-jadval. Yozuvlarning tasnifi

Yozuvlarni taqqoslash	Holati
$(a3, b3)$	Moslik bor
$(a4, b4)$	Moslik yo'q

Yuqoridagi bosqichlardan ko'rish mumkinki dastlab mos kelmaydigan natijalarni olib tashlaymiz, chunki biz faqat moslik bo'ladigan holatlar muhim. Ikkinchidan, oldin moslashtirilgan maydonlarni olib tashlaymiz chunki yangi maydonlar bilan ishlashni osonlashadi. Masalan ma'lumotlar_bazasi_1 dagi bitta yozuv

ma'lumotlar_bazasi_2dagi ikkita yozuvlar bilan mos keladigan bo'lsa, bunda mos kelish ehtimoli past bo'lgan juftlikni olib tashlaymiz. Lekin bu amal keyinchalik salbiy natijalar keltirib chiqarishi mumkin. Ikki yoki undan ortiq juftlik bir xil yuqori ehtimollikka ega bo'lsa, unda bu juftliklarni saqlab turamiz. Agar zarurat tug'ilsa faqat bittasi moslik aniqlangan taqdirda ham moslashuvni amalga oshirish uchun ikkala yozuvni ham saqlashni tanlashimizga to'g'ri keladi [12,13].

IV. XULOSA

Tadqiqot davomida turli ma'lumotlar bazasidagi yozuvlarni moslashtirishda bloklash usulining samaradorligi va o'rni boshqa usullarga yaxshi ekanligi tajribalar asosida isbotlandi. Yozuvlarni bog'lash jarayoni qanday amalga oshirilishda bloklash jarayoning muhim

bosqichlari keltirib o'tildi. Bu esa turli ma'lumotlar bazalaridagi yozuvlarni moslashtirish uchun juda muhim bosqich ekanligini ko'rsatildi. Bunda indekslash qadami uchun standart bloklash tanlandi, chunki u tezroq (blokirovka kalitlari qoliplarini tanlash shart emas). Nazorat ostida o'qitishda ma'lumotlar bazasida aniqlangan tanlanma asosida ma'lumotlar bazasida yozuvlar mos va mos kelmaydigan juftliklardan qarorlar daraxtlari, logistik regressiya va boshqa chuqur o'qitishlar orqali natijalarni shakllantirish yondashuvlari keltirib o'tildi.

ADABIYOTLAR

- [1] Sun'iy intellekt texnologiyalarini 2030-yilga qadar rivojlantirish strategiyasini tasdiqlash to'g'risida. O'zbekiston Respublikasi Prezidentining qarori, 14.10.2024 yildagi PQ-358-son
- [2] «O'zbekiston — 2030» strategiyasi to'g'risida. O'zbekiston Respublikasi Prezidentining Farmoni, 11.09.2023 yildagi PF-158-son
- [3] *I. P. Fellegi and A. B. Sunter*, "A Theory for Record Linkage," *Journal of the American Statistical Association*, vol. 64, no. 328, p. 1183, 12 1969.
- [4] *S. Yan, D. Lee, M.-Y. Kan, and L. C. Giles*, "Adaptive Sorted Neighborhood Methods for Efficient Record Linkage," *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, p. 185–194, 2007. <http://doi.acm.org/10.1145/1255175.125523>
- [5] *P. Christen*, "A Survey of Indexing Techniques for Scalable Record Linkage and Deduplication," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 9, pp. 1537–1555, 9 2012. <http://ieeexplore.ieee.org/document/5887335>
- [6] *P. Christen*, *Data Matching*. Springer Berlin Heidelberg, 2012, no. Chapter 1. [Online]. Available: <http://dx.doi.org/10.1007/978-3-662-10876-5>
- [7] *Nurmamatov, M., Kulmirzayeva, Z.* "Development of an Intelligent System for Optimization of Employment Information Using Genetic Algorithms" *AIP Conference Proceedings*, 2024, 3147(1), 040006. <https://doi.org/10.1063/5.0210279>
- [8] *Akhatov A., Nurmamatov M., Nazarov F.* "Intelligent modeling and optimization of processes in the labour market" *Artificial Intelligence, Blockchain, Computing and Security - Proceedings of the International Conference on Artificial Intelligence, Blockchain, Computing and Security, ICABCS 2023, 2024, 2, страницы 694–699*
- [9] *AA Rustamovich, N Mekhriddin, N Fayzullo, N Sabharwal* "Intelligent system of labor market regulation based on the evolutionary modeling of employment" *Proceedings - 2022 4th International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2022, 2022, c. 2534–2539.*
- [10] *Nurmamatov M.Q.*, Modern methods of increasing the efficiency of the labor market. // "ILM-FAN" *electron journali. UZA.* –b. 373-383. 2022. <http://uza.uz/posts/363383>.
- [11] *Akhatov A.R., Nurmamatov M.Q., Mardonov D.R., Nazarov F.M.* Improvement of mathematical models of the rating point system of employment // *Scientific journal Samarkand state university.* 2021. – №1(125). –P. 100-107.
- [12] *Nurmamatov M.Q., Sariyev Sh.N.*, Genetik algoritmlar asosida turli sinfli ma'lumotlarni o'zaro moslashtirish algoritmlari. *Sh.Rashidov nomidagi Samarqand Davlat Universiteti Ilmiy axborotnomasi.* 3-son (145/1) aniq va tabiiy fanlar yo'nalishi. 77-83.
- [13] *Axatov A.R., Nurmamatov M.Q., Nazarov F.M.* 2022. "Mathematical Models of Coordination of Population Employment in the Labor Market" // *Ra journal of applied research. India / –Vol. 8, Issue 2. P 111–119.* doi:<https://doi.org/10.47191/rajar/v8i2.09>
- [14] *Akhatov A.R., Nurmamatov M.Q., Mardonov D.* 2020. "Mathematical models of the process of monitoring the social status and employment of the population", *Scientific and technical journal of the Fergana Polytechnic Institute.* -Volume 24, No. 5. - pp. 150–157.

Поступила в редакцию 10.09.2024

Citation: *Nurmamatov M.Q.* (2024). Ma'lumotlarni moslashtirishning zamonaviy yondashuvlari. *Raqamli texnologiyalarning nazariy va amaliy masalalari xalqaro jurnali.* 7(4). – B. 20-26. <https://doi.org/10.62132/ijdt.v7i4.215>

MODERN APPROACHES TO DATA MATCHING

Nurmamatov M.K.¹

¹ Samarkand State University named after Sharof Rashidov, Samarkand, Uzbekistan
mehriddinnur@gmail.com

Abstract. *Today, the volume of information is growing at a rapid pace and the process of giving opinions for each person using records from different databases is improving. By aggregating records from different databases into a single database by matching them, decision-making has the potential to optimize workflows as well as achieve cost-effectiveness. Such adjustments in Medicine, Online Sales, Employment Lists improve time and economic performance. For this reason, the methods, models and approaches of matching records in different databases are explored in this article.*

Keywords: *K-means, machine learning, employment, database, blocking, Q-gram, expert assessment.*

СОВРЕМЕННЫЕ ПОДХОДЫ АДАПТАЦИИ ДАННЫХ

Нурмаматов М.К.¹

¹ Самаркандский государственный университет имени Шарофа Рашидова,
Самарканд, Узбекистан
mehriddinnur@gmail.com

Аннотация. *Сегодня объем информации растет в виде четких изображений, а процесс составления выводов на основе записей в разных базах данных по каждому человеку совершенствуется. Принятие решений путем объединения записей из разных баз данных в единую базу путем адаптации записей в разных базах данных обеспечивает оптимизацию рабочих процессов и экономичность. Подобные изменения в системах здравоохранения, онлайн-продаж и занятости позволяют сократить время и улучшить экономические показатели. По этой причине в данной статье рассматриваются методы, модели и подходы для сопоставления записей в разных базах данных.*

Ключевые слова: *K-means, машинное обучение, трудоустройство, база данных, блокировка, Q-грамма, экспертная оценка.*