ARALASH TURDAGI BELGILAR BO‘YICHA O‘XSHASHLIK VA BO‘SAG‘AVIY QIYMATLARNI SHAKLLANTIRISH

Ro‘zibayev Ortiq Baxtiyorovich

Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti
E-mail:o.ruzibaev@tuit.uz

K E Y W O R D S

multisinf, belgi, pretsedent obyekt, o‘xshashlik indikatori, adaptiv bo‘sag‘a, aralash belgilar, binar moslik vektori, shovqin, tasniflash algoritmi

A B S T R A C T

Ushbu maqolada timsollarni tanib olish va informativ belgilarni tanlash masalasining muhimligi hamda uni yechishda pretsedent ob’ektlarni tanlashning ahamiyati yoritilgan. Tadqiqot doirasida nominal va qiymatli belgilar fazosida shakllantirilgan o‘quv tanlanmasi tahlil qilindi. Multisinf obyektlari o‘rtasidagi o‘xshashlikni baholash va sinf ichidagi pretsedentlarning o‘zaro farqlanish darajasini aniqlashga asoslangan mezonlar ishlab chiqildi. Aralash turdagi belgilarni qayta ishlash uchun nominal belgilar bo‘yicha tenglikka asoslangan, sonli belgilar uchun esa adaptiv bo‘sag‘alar yordamida shakllantirilgan o‘xshashlik indikatorlari taklif qilindi. Natijada turli belgilar yagona binar moslik vektorida ifodalash imkoniyati yaratildi. Tadqiqotda sonli belgilar orasidagi farq qiymatini bevosita bo‘sag‘a sifatida qo‘llash natijasida yuzaga keladigan noaniqliklar bartaraf etilib, u belgilarning ajralish masofasi sifatida talqin qilindi. Har bir multisinfning ichki xususiyatlarini hisobga olgan holda pretsedentlararo o‘xshashlikni aniqlash usuli ishlab chiqildi.

KIRISH

Ob’ektlarni pretsedentlar bilan taqqoslashda barcha belgilarni yagona qoida orqali baholash mumkin emas. Nominal belgida qiymatlar orasida arifmetik masofa ma’noga ega emas; sonli belgida esa mutlaq tenglik haddan tashqari qat’iy shart hisoblanadi. Masalan, «shish mavjud/mavjud emas» belgisi uchun tenglik to‘g‘ridan-to‘g‘ri qo‘llansa, tana harorati 37,6 va 37,7 °C bo‘lgan ikki bemorni farqli deb hisoblash klinik jihatdan asossiz. Shuning uchun o‘xshashlik har bir belgi turiga mos lokal indikatorlar orqali aniqlanadi [5, 7, 14,15].

ASOSIY QISM

X_{pq} multisinf pretsedenti $\hat{x}_{pq}$ va uning x_{pqk} ob’ektining j-belgi bo‘yicha o‘xshashligini

$$\rho_{pq}^{j}(\hat{x}_{pq}, x_{pqk}) \in \{0,1\}, \qquad j = \overline{1,N}, \qquad (1)$$

orqali belgilaymiz. Agar $j \in J_{nom}$ bo‘lsa, nominal o‘xshashlik quyidagicha aniqlanadi:

$$\rho_{pq}^{j}(\hat{x}_{pq}, x_{pqk}) = \begin{cases} 1, & \hat{x}_{pq}^{j} = x_{pqk}^{j}, \\ 0, & \hat{x}_{pq}^{j} \neq x_{pqk}^{j}. \end{cases} \qquad (2)$$

Ushbu qoida binar, nominal va kategoriyali belgilar uchun tabiiy. Agar ayrim kategoriyalar klinik jihatdan yaqin bo‘lsa, masalan, kasallik bosqichining qo‘shni darajalari, $0/1$ indikator o‘rniga ekspertlik o‘xshashlik matritsasi $G_j = (g_{ab}^{j})$ qo‘llanishi mumkin. Bunda

$$\rho_{pq}^{j}(\hat{x}_{pq}, x_{pqk}) = g^{j}_{\hat{x}_{pq}^{j}, x_{pqk}^{j}}, \qquad 0 \le g_{ab}^{j} \le 1, \qquad (3)$$

bo‘ladi. Ammo asosiy algoritmning aniq va sodda talqinini saqlash maqsadida bazaviy holda (2) binar tenglik qoidasi qabul qilinadi.

(cc) BY

Agar $j \in J_{num}$ bo‘lsa, sonli belgilar uchun pretsedentdan og‘ish qabul qilinadigan ε_{pq}^{j} bo‘sag‘a bilan taqqoslanadi:

$$\rho_{pq}^{j}(\hat{x}_{pq}, x_{pqk}) = \begin{cases} 1, & |\hat{x}_{pq}^{j} - x_{pqk}^{j}| \le \varepsilon_{pq}^{j}, \\ 0, & |\hat{x}_{pq}^{j} - x_{pqk}^{j}| > \varepsilon_{pq}^{j}. \end{cases} \quad (4)$$

Bo‘sag‘a multisinfning ichki tarqalishini hisobga olishi lozim. Yuklangan belgilashlar tizimiga muvofiq, pretsedentdan qolgan ob’ektlargacha bo‘lgan mutlaq og‘ishlarning o‘rtacha qiymati qabul qilinadi:

$$\varepsilon_{pq}^{j} = \frac{1}{m_{pq}-1} \sum_{\substack{k=1 \\ x_{pqk} \neq \hat{x}_{pq}}}^{m_{pq}} |\hat{x}_{pq}^{j} - x_{pqk}^{j}|, \qquad j \in J_{num}. \quad (5)$$

Ushbu yozuv dastlabki formulaning mazmunini aniqlashtiradi: yig‘indida pretsedentning o‘zi qatnashmaydi, shuning uchun maxraj $m_{pq} - 1$ ga teng. Agar pretsedent x_{pqs} indeksga ega bo‘lsa, yig‘indi $k \neq s$ shart bilan yoziladi. (5) bo‘sag‘a multisinf ichidagi tabiiy variatsiyani avtomatik hisobga oladi: bir jinsli guruhda u kichik, tarqoq guruhda kattaroq bo‘ladi.

O‘rtacha mutlaq og‘ish shovqinli yoki keskin chet ob’ektlarga sezgir bo‘lishi mumkin. Tibbiy ma’lumotda laborator xatolar, noto‘g‘ri kiritilgan qiymatlar yoki kam uchraydigan og‘ir holatlar bo‘sag‘ani sun’iy kattalashtirishi ehtimol. Shu bois robast variant sifatida mediana mutlaq og‘ishi qo‘llanishi mumkin:

$$\tilde{\varepsilon}_{pq}^{j} = c_{MAD}\, median_{x_{pqk} \neq \hat{x}_{pq}} |\hat{x}_{pq}^{j} - x_{pqk}^{j}|, \quad (6)$$

bu yerda c_{MAD} – shkalalash koeffitsienti. Bazaviy algoritmda (5), shovqinga chidamli modifikatsiyada esa (6) ishlatiladi. Qaysi variantni tanlash ma’lumot taqsimoti va klinik tekshiruv natijasiga bog‘liq.

Belgilar diapazoni turlicha bo‘lganida mutlaq og‘ishni normallashtirish kerak. $R_j = x_{max}^{j} - x_{min}^{j}$ – j-belgining o‘quv qismdagi diapazoni bo‘lsa,

$$d_{pqk}^{j} = \frac{|\hat{x}_{pq}^{j} - x_{pqk}^{j}|}{R_j + \xi}, \quad (7)$$

normallashgan og‘ish kiritiladi; $\xi > 0$ nolga bo‘lishdan himoya qiladi. Shunda bo‘sag‘a ham normallashgan shkalada hisoblanishi mumkin. Biroq klinik birlikdagi farq tushunarli va o‘lchovlar bir xil bo‘lsa, (5) shakl talqin qilish uchun qulayroq.

Ba’zi belgilarda klinik referens oraliqlari ma’lum. Masalan, laborator ko‘rsatkich uchun shifokor qabul qiladigan farq τ_j^{clin} belgilangan bo‘lsa, statistik va klinik bo‘sag‘ani birlashtirish mumkin:

$$\varepsilon_{pq}^{j,comb} = \omega \varepsilon_{pq}^{j} + (1-\omega)\tau_j^{clin}, \qquad 0 \le \omega \le 1. \quad (8)$$

yoki ehtiyotkor qoida sifatida $\min(\varepsilon_{pq}^{j}, \tau_j^{clin})$ qabul qilinadi. Bu statistik tarqalish juda katta bo‘lganda klinik jihatdan nomaqbul o‘xshashlikka yo‘l qo‘ymaslikka xizmat qiladi.

(2) va (4) asosida pretsedent–ob’ekt o‘xshashlik vektori shakllanadi:

$$\rho_{pq}(\hat{x}_{pq}, x_{pqk}) = \left(\rho_{pq}^{1}(\hat{x}_{pq}, x_{pqk}), \rho_{pq}^{2}(\hat{x}_{pq}, x_{pqk}), \dots, \rho_{pq}^{N}(\hat{x}_{pq}, x_{pqk})\right). \quad (9)$$

Vektorning har bir komponenti muayyan belgi bo‘yicha moslikni anglatadi. Mos kelgan belgilar soni

$$\varkappa(\hat{x}_{pq}, x_{pqk}) = \sum_{j=1}^{N} \rho_{pq}^{j}(\hat{x}_{pq}, x_{pqk}) \quad (10)$$

orqali hisoblanadi. Shunda ob’ektlarning umumiy o‘xshashlik darajasi

$$\nu(\hat{x}_{pq}, x_{pqk}) = \frac{\varkappa(\hat{x}_{pq}, x_{pqk})}{N} \cdot 100\%. \quad (11)$$

(11) ko‘rsatkich barcha belgilar teng ahamiyatli va to‘liq mavjud bo‘lgan bazaviy holatga mos. Yetishmaydigan qiymatlar mavjud bo‘lsa, maxraj faqat taqqoslangan belgilar soniga teng bo‘lishi lozim:

$$\nu_w(\hat{x}_{pq}, x_{pqk}) = \frac{\sum_{j=1}^{N} w_{pq}^{j} w_{pqk}^{j} \rho_{pq}^{j}(\hat{x}_{pq}, x_{pqk})}{\sum_{j=1}^{N} w_{pq}^{j} w_{pqk}^{j}} \cdot 100\%. \quad (12)$$

(cc) BY

Agar umumiy mavjud belgilar soni oldindan belgilangan N_{min} dan kam bo‘lsa, o‘xshashlik ishonchsiz deb belgilanadi. Bu qoida juda kam ma’lumotli bemorni tasodifan yuqori o‘xshash deb qabul qilishning oldini oladi. Informativ belgilar tanlangan holda o‘xshashlik faqat $\lambda_j = 1$ bo‘lgan komponentlar bo‘yicha hisoblanadi:

$$\nu_\lambda(\hat{x}_{pq}, x_{pqk}) = \frac{\sum_{j=1}^{N} \lambda_j \rho_{pq}^{j}(\hat{x}_{pq}, x_{pqk})}{\sum_{j=1}^{N} \lambda_j} \cdot 100\% = \frac{100}{\ell} \sum_{j=1}^{N} \lambda_j \, \rho_{pq}^{j}(\hat{x}_{pq}, x_{pqk}). \quad (13)$$

Yetishmaydigan qiymatlar bilan birgalikda esa

$$\nu_{\lambda,w}(\hat{x}_{pq}, x_{pqk}) = \frac{\sum_{j=1}^{N} \lambda_j w_{pq}^{j} w_{pqk}^{j} \rho_{pq}^{j}(\hat{x}_{pq}, x_{pqk})}{\sum_{j=1}^{N} \lambda_j w_{pq}^{j} w_{pqk}^{j}} \cdot 100\%. \quad (14)$$

(13) formula informativ belgilarni tanlash masalasining markaziy bog‘lanishini ko‘rsatadi. Belgi qism to‘plamdan chiqarilganda u na o‘xshashlikni, na maxrajni shakllantiradi. Demak, qism to‘plamni o‘zgartirish multisinflarning geometrik tuzilishi va pretsedentlarning qamrovini bevosita o‘zgartiradi.

Belgilar ahamiyati turlicha bo‘lsa, binar tanlashga qo‘shimcha ravishda $\theta_j \geq 0$ vazn kiritish mumkin:

$$\nu_{\lambda,\theta}(\hat{x}_{pq}, x_{pqk}) = \frac{\sum_{j=1}^{N} \lambda_j \theta_j \rho_{pq}^{j}(\hat{x}_{pq}, x_{pqk})}{\sum_{j=1}^{N} \lambda_j \theta_j} \cdot 100\%. \quad (15)$$

Biroq dissertatsiyaning asosiy vazifasi aynan belgilar qism to‘plamini tanlash bo‘lgani uchun, vaznlar yordamchi modifikatsiya sifatida qaraladi. Binar λ natijani klinik talqin qilishni osonlashtiradi: har bir belgi yo tanlangan, yo tanlanmagan.

Turli multisinf pretsedentlari orasidagi sonli belgi bo‘yicha farq yuklangan belgilashga muvofiq

$$\varepsilon_{pi,pt}^{j} = \left|\hat{x}_{pi}^{j} - \hat{x}_{pt}^{j}\right|, \qquad i \neq t, \quad (16)$$

orqali belgilanadi. Mazkur kattalikni to‘g‘ridan-to‘g‘ri «o‘xshashlik bo‘sag‘asi» sifatida ishlatish mantiqiy tavtologiyaga olib kelishi mumkin: farq o‘z qiymati bilan taqqoslansa, shart har doim bajariladi. Shu sababli (16) ni pretsedentlararo belgi masofasi yoki ajralish qiymati sifatida talqin qilish matematik jihatdan to‘g‘riroq. Pretsedentlarning j–belgi bo‘yicha o‘xshashlik indikatori esa ularning ichki bo‘sag‘alaridan kelib chiqib aniqlanadi:

$$\rho_{p,it}^{j}(\hat{x}_{pi}, \hat{x}_{pt}) = \begin{cases} 1, & j \in J_{nom},\ \hat{x}_{pi}^{j} = \hat{x}_{pt}^{j}, \\ 1, & j \in J_{num},\ \left|\hat{x}_{pi}^{j} - \hat{x}_{pt}^{j}\right| \leq min(\varepsilon_{pi}^{j}, \varepsilon_{pt}^{j}), \\ 0, & aks\ holda. \end{cases} \quad (17)$$

(17) qoidasi quyidagi mazmunga ega: agar ikki pretsedent orasidagi farq har ikkala multisinfning ichki qabul qilinadigan variatsiyasidan ham kichik bo‘lsa, ular mazkur belgi bo‘yicha o‘xshash hisoblanadi. Agar farq kamida bir guruhning ichki chegarasidan katta bo‘lsa, belgi ajratuvchi hisoblanadi. Nominal belgilarda oddiy tenglik saqlanadi.

Pretsedentlararo umumiy o‘xshashlik

$$\nu_\lambda(\hat{x}_{pi}, \hat{x}_{pt}) = \frac{100}{\ell} \sum_{j=1}^{N} \lambda_j \, \rho_{p,it}^{j}(\hat{x}_{pi}, \hat{x}_{pt}) \quad (18)$$

orqali aniqlanadi. Shunda asosiy pretsedent shartlari tanlangan fazoda

$$\nu_\lambda(\hat{x}_{pq}, x_{pqk}) \geq \delta, \qquad \nu_\lambda(\hat{x}_{pi}, \hat{x}_{pt}) \leq \frac{\delta}{3} \quad (19)$$

ko‘rinishiga keladi.

O‘xshashlikning binar komponentlarga asoslanishi natijani tushunarli qiladi, ammo chegara atrofida keskin sakrash paydo qiladi. Masalan, farq ε dan juda oz kichik bo‘lsa $\rho = 1$, juda oz katta bo‘lsa $\rho = 0$. Bu muammoni yumshatish uchun sonli belgilarda uzluksiz o‘xshashlik funksiyasi kiritilishi mumkin:

$$s_{pq}^{j}(\hat{x}_{pq}, x_{pqk}) = exp\left(-\frac{\left|\hat{x}_{pq}^{j} - x_{pqk}^{j}\right|}{\varepsilon_{pq}^{j} + \xi}\right), \qquad j \in J_{num}. \quad (20)$$

Yoki chiziqli funksiya

$$s_{pq}^{j} = max\left\{0{,}1 - \frac{\left|\hat{x}_{pq}^{j} - x_{pqk}^{j}\right|}{\varepsilon_{pq}^{j} + \xi}\right\}. \quad (21)$$

(cc) BY

Biroq binar sxema pretsedentni tanlash shartlarini va «nechta belgi mos keldi?» degan klinik tushuntirishni aniq saqlaydi. Shu sababli asosiy algoritmda (2), (4) va (17), ehtimoliy yoki yumshoq modifikatsiyada esa (20)–(21) qo‘llanishi mumkin.

δ bo‘sag‘ani tanlash model sezgirligi va spesifikligiga bevosita ta’sir qiladi. δ juda past bo‘lsa, bir pretsedent turli fenotip ob’ektlarini ham qoplaydi va sinflar orasidagi farq yo‘qoladi. δ juda yuqori bo‘lsa, har bir multisinf uchun pretsedent topish qiyinlashadi, yangi ob’ektlar ko‘pincha «noaniq» deb belgilanadi. Shuning uchun δ quyidagi manbalar asosida baholanadi:

- klinik ekspert tomonidan belgilangan minimal moslik;
- o‘quv tanlanmada ichki qamrov va tashqi xatoning muvozanati;
- nested cross-validationda balanced accuracy, sensitivity yoki utility funksiyasini maksimallashtirish;
- pretsedentlar soni va «noaniq» qarorlar ulushini cheklash.

Formal ravishda

$$\delta^* = \underset{\delta \in \Delta}{argmax}\{BA(\delta) - \beta_1 Reject(\delta) - \beta_2 Complexity(\delta)\} \quad (22)$$

ko‘rinishida tanlanishi mumkin. Bu yerda Reject – noaniq ob’ektlar ulushi, Complexity – umumiy pretsedentlar yoki multisinflar soni, β_1, β_2 – ahamiyat koeffitsientlari.

O‘xshashlik funksiyasining metrik xususiyatlari ham e’tiborga olinadi. ν o‘xshashlik bo‘lib, masofa emas; undan

$$d_\lambda(u, v) = 1 - \frac{\nu_\lambda(u,v)}{100} \quad (23)$$

masofasimon kattalik hosil qilish mumkin. Nominal va bo‘sag‘alangan sonli komponentlarda simmetriklik ta’minlanadi, ammo uchburchak tengsizligi har doim bajarilmasligi mumkin. Shu sababli ushbu kattalik «dissimilyarlik» sifatida talqin qilinadi. Klasterlash algoritmi qat’iy metrika talab qilsa, Gower masofasi yoki maxsus aralash masofadan foydalanib, pretsedent tanlashda esa (11)–(18) moslik mezonlari saqlanadi [7].

Belgilar bo‘yicha o‘xshashlikni hisoblash natijasida har bir ob’ekt uchun tushuntirish profili hosil bo‘ladi. Masalan,

$$E(x; \hat{x}_{pq}) = \{j: \lambda_j = 1, \rho_{pq}^j(\hat{x}_{pq}, x) = 0\} \quad (24)$$

to‘plam bemorning eng yaqin pretsedentdan farq qilgan informativ belgilarini ko‘rsatadi. Tibbiy qarorni qo‘llab-quvvatlash tizimida ushbu to‘plam «mos kelmagan belgilari» sifatida chiqarilishi, shifokorga qo‘shimcha tekshiruv yo‘nalishini ko‘rsatishi mumkin. Agar farqlangan belgilar klinik jihatdan xavfli bo‘lsa, umumiy o‘xshashlik yuqori bo‘lsa ham, maxsus qoida bilan qaror qayta ko‘rib chiqiladi.

Shunday qilib, qabul qilingan o‘xshashlik modeli uchta muhim xususiyatni ta’minlaydi. Birinchidan, nominal va sonli belgilar yagona $[0,1]$ moslik shkalasiga keltiriladi. Ikkinchidan, sonli belgilar uchun bo‘sag‘alar har bir multisinfning ichki tarqalishiga moslashadi. Uchinchidan, tanlash vektori λ bilan bog‘lanish belgilar qism to‘plamning pretsedent qamrovi va ajraluvchanligiga ta’sirini bevosita baholash imkonini beradi. Endi mazkur shartlarga asoslangan optimal belgilar to‘plami masalasini shakllantirish mumkin.

XULOSA

Pretsedentlarni tanblash uchun ikki asosiy shart aniqlandi: multisinf ob’ektlari bilan o‘xshashlik $\nu(\hat{x}_{pq}, x_{pqk}) \geq \delta$ bo‘lishi va bir sinfning turli multisinf pretsedentlari o‘xshashligi $\nu(\hat{x}_{pi}, \hat{x}_{pt}) \leq \delta/3$ chegaradan oshmasligi lozim.

Nominal belgilar uchun tenglikka asoslangan, sonli belgilar uchun esa multisinf ichidagi o‘rtacha mutlaq og‘ishdan hosil qilingan ε_{pq}^j adaptiv bo‘sag‘aga asoslangan ρ_{pq}^j o‘xshashlik indikatorlari ishlab chiqildi. Natijada aralash belgilar yagona binar moslik vektorida ifodalandi.

Pretsedentlar orasidagi $\varepsilon_{pi,pt}^j = \left|\hat{x}_{pi}^j - \hat{x}_{pt}^j\right|$ kattalikni to‘g‘ridan-to‘g‘ri bo‘sag‘a sifatida ishlatishda yuzaga keladigan mantiqiy noaniqlik bartaraf etilib, u belgi bo‘yicha ajralish masofasi sifatida talqin qilindi; pretsedentlararo o‘xshashlik esa har ikki multisinfning ichki bo‘sag‘alari asosida aniqlandi.

(cc) BY

Har bir sinfdagi president ob’ektlarni aniq mezonlar asosida tanlash, ya’ni o‘zaro o‘xshashlik va farq darajalarini hisobga olish, tasniflash jarayonida yuzaga keladigan noaniqliklarni kamaytiradi. Shu bilan birga, algoritmning ko‘p maqsadli, masshtablashuvchan va domenga moslashuvchan xususiyatlari uni boshqa sohalarda, jumladan bioinformatika, marketing, ijtimoiy tahlil va sanoatdagi monitoring tizimlarida ham qo‘llash imkonini yaratadi.

Kelgusidagi tadqiqotlarda taklif etilgan algoritmni real tibbiy ma’lumotlar asosida sinovdan o‘tkazish, shuningdek, president ob’ektlarni avtomatik tanlashni chuqur o‘rganish orqali uning intellektual imkoniyatlarini yanada kengaytirish maqsadga muvofiq hisoblanadi.

ADABIYOTLAR

1. *Журавлев Ю.И.* Об алгебраическом подходе к решению задач распознавания или классификации.// Проблемы кибернетики, 1978. Вып. 33. С. 5-68.
2. *Неймарка Ю.* Распознавание образов и медицинская диагностика/ Под ред. М.: Наука, 1972.
3. *Енюков И.С.* Методы, алгоритмы, программы многомерного статистического анализа. - М.: Финансы и статистика, 1986.
4. *Guyon I., Elisseeff A.* An Introduction to Variable and Feature Selection // Journal of Machine Learning Research. – 2003. – Vol. 3. – P. 1157–1182.
5. *Varoquaux G., Cheplygina V.* Machine Learning for Medical Imaging: Methodological Failures and Recommendations for the Future // npj Digital Medicine. – 2022. – Vol. 5. – Art. 48. DOI: 10.1038/s41746-022-00592-y.
6. *Collins G.S., Moons K.G.M.,* Dhiman P. et al. TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods // BMJ. – 2024. – Vol. 385. – e078378. DOI: 10.1136/bmj-2023-078378.
7. *Vasey B., Nagendran M., Campbell B.* et al. Reporting Guideline for the Early-Stage Clinical Evaluation of Decision Support Systems Driven by Artificial Intelligence: DECIDE-AI // Nature Medicine. – 2022. – Vol. 28. – P. 924–933. DOI: 10.1038/s41591-022-01772-9.
8. *Liu X., Rivera S.C., Moher D.* et al. Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension // Nature Medicine. – 2020. – Vol. 26. – P. 1364–1374. DOI: 10.1038/s41591-020-1034-x.
9. *Chicco D., Jurman G.* The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation // BMC Genomics. – 2020. – Vol. 21. – Art. 6. DOI: 10.1186/s12864-019-6413-7.
10. *Tibshirani R.* Regression Shrinkage and Selection via the Lasso // Journal of the Royal Statistical Society: Series B. – 1996. – Vol. 58, No. 1. – P. 267–288. DOI: 10.1111/j.2517-6161.1996.tb02080.x.
11. *Cover T.M., Hart P.E.* Nearest Neighbor Pattern Classification // IEEE Transactions on Information Theory. – 1967. – Vol. 13, No. 1. – P. 21–27. DOI: 10.1109/TIT.1967.1053964.
12. *Huang Z.* Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values // Data Mining and Knowledge Discovery. – 1998. – Vol. 2. – P. 283–304. DOI: 10.1023/A:1009769707641.
13. *Kapoor S., Narayanan A.* Leakage and the Reproducibility Crisis in Machine-Learning-Based Science // Patterns. – 2023. – Vol. 4, No. 9. – Art. 100804. DOI: 10.1016/j.patter.2023.100804.
14. *Van Calster B., McLernon D.J., van Smeden M., Wynants L., Steyerberg E.W.* Calibration: The Achilles Heel of Predictive Analytics // BMC Medicine. – 2019. – Vol. 17. – Art. 230. DOI: 10.1186/s12916-019-1466-7.
15. *Нишанов А.Х., Жўраев Ғ.П., Нарзиев Н.Б.,* Хушвақтов Ж. Тимсолларни таниб олишда ε–бўсақавий қийматни танлаш алгоритми// «Транспортда инновацион ахборот технологияларини ривожлантиришнинг долзарб масалалари» I–Республика илмий–техник анжуманининг материаллари. – Тошкент, 2021. 168 – 172 б.
16. *Lundberg S.M., Lee S.-I.* A Unified Approach to Interpreting Model Predictions //

(cc) BY

Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 4765–4774.
17. *Wolff R.F., Moons K.G.M., Riley R.D.* et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies // Annals of Internal Medicine. – 2019. – Vol. 170, No. 1. – P. 51–58. DOI: 10.7326/M18-1376.
18. *Saito T., Rehmsmeier M.* The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets // PLoS ONE. – 2015. – Vol. 10, No. 3. – e0118432. DOI: 10.1371/journal.pone.0118432.
19. *Nishanov A.Kh., Rakhmanov A.T., Ruzibaev O.B., Allamov O.T., Zaripov F.M.* Алгоритм многоклассовой классификации с использованием шарового покрытия. Высокопроизводительные вычислительные системы и технологии в научных исследованиях, автоматизации управления и производства: XII Междунар. науч.-практ. конф. (ВВСТ-2022), 20–21 мая 2022 г. // Высокопроизводительные вычислительные системы и технологии. — 2022. — Т. 6, № 1. — С. 119–125.
20. *Zhuravlev Y.I., Ryazanov V.V.* Solution of instance-based recognition problems with a large number of classes. Doklady Mathematics. 2017. Vol. 96, No. 2. P. 488–490.
21. *Zhuravlev Y.I.* On the algebraic approach to solving problems of recognition or classification. Problems of Cybernetics. 1978. Vol. 33. P. 5–68.
22. *Gorelik A.L.* General statement of the problem of object and phenomenon recognition. Cybernetics. 1980. No. 6. P. 72–75.
23. *Nishanov A.Kh., Akbaraliev B.B., Samandarov B.S., Akhmedov O.K., Tajibaev S.K.* An algorithm for classification, localization and selection of informative features in the space of polytypic data. Webology. 2020. Vol. 17, No. 1. P. 341–364.
24. *Rakhmanov A.T., Akbaraliev B.B., Ruzibaev O.B.* On one modified method of solving the classification problem. Chemical Technology. Control and Management. 2014. No. 4. P. 85–91.
25. *Rakhmanov A.T., Ruziboev O.B.* To the solution of the classification problem for three classes. Science and World. 2015. Vol. 2, No. 18. P. 21–25.
26. *Nishanov A.X., Ruzibaev O.B., Tran N.H.* Modification of decision rules “Ball Apolonia” in the problem of classification. International Conference on Information Science and Communications Technologies (ICISCT). 2016. P. 1–3.
27. *Nishanov A.Kh., Ruzibaev O.B., Chedjou J.C., Kyamakya K.* et al. Algorithm for the selection of informative symptoms in the classification of medical data. Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference. 2020. P. 647–658.
28. *Chandrashekar, G., & Sahin, F.* (2014). A Survey on Feature Selection Methods. Computers & Electrical Engineering, 40(1), 16–28.
29. *Li, J., Cheng, K., Wang, S.,* et al. (2017). Feature Selection: A Data Perspective. ACM Computing Surveys, 50(6), 94.
30. *Saeys, Y., Inza, I., & Larrañaga, P.* (2007). A Review of Feature Selection Techniques in Bioinformatics. Bioinformatics, 23(19), 2507–2517.
31. *Guyon, I., Weston, J., Barnhill, S., & Vapnik, V.* (2002). Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning, 46(1–3), 389–422.
32. *Remeseiro, B., & Bolon-Canedo, V.* (2019). A Review of Feature Selection Methods in Medical Applications. Computers in Biology and Medicine, 112, 103375.
33. *Al Mamun, A., Sohel, F.,* et al. (2021). A Comprehensive Review of Feature Selection Methods for Biomedical Data. IEEE Access, 9, 114125–114148.

(cc) BY *dtai.uz*