AXBOROT TIZIMLARINING MA'LUMOTLARINI INTELLEKTUAL TAHLIL QILISHDA YETISHMAYOTGAN (NAN) QIYMATLARNI QAYTA ISHLASH ALGORITMLARINING NAZARIY TAHLILI

Orifov Oxunjon Fazliddinzoda

Sharof Rashidov nomidagi Samarqand davlat universiteti 1-bosqich magistranti.

E-mail: *okhunjonorifov555@gmail.com*

https://doi.org/10.5281/zenodo.15186061

Annotatsiya. *Ushbu maqolada axborot tizimlarida uchraydigan yetishmayotgan (NaN) qiymatlarni aniqlash va ularni to'ldirishga oid nazariy yondashuvlar chuqur tahlil qilinadi.*

Ma'lumotlar to'plamining sifati intellektual tahlil jarayonlarining ishonchliligi va samaradorligiga bevosita ta'sir ko'rsatadi. Shu bois, NaN qiymatlarni bartaraf etish algoritmlari hozirgi davr axborot texnologiyalarida dolzarb masala sifatida namoyon bo'lmoqda. Maqolada 7 ta asosiy metod – o'rtacha qiymat, median, moda, regressiya, K eng yaqin qo'shni (KNN), interpolatsiya va sun'iy neyron tarmoqlar yordamida to'ldirish – nazariy jihatdan o'rganildi. Har bir metod hisoblash murakkabligi, dispersiyaga ta'siri, modelga sezgirlik darajasi, bashorat aniqligi va ma'lumotlar tiplariga nisbatan moslashuvchanlik mezonlari bo'yicha baholandi. Tadqiqotda metodlarning kuchli va zaif jihatlari asoslangan holda tahlil qilinib, turli kontekstlar uchun optimal yondashuvni tanlash zarurati asoslab berildi. Shuningdek, kombinatsiyalangan (hibrid) algoritmlar yordamida kompleks yondashuvlar ishlab chiqish istiqbollari ham ilgari surildi. Ushbu maqola axborot tizimlarida intellektual tahlil sifatini oshirish, sun'iy intellekt modellarining aniqligini ta'minlash hamda raqamli transformatsiya jarayonlarini samarali tashkil etish uchun muhim nazariy asos bo'lib xizmat qiladi.

Kalit so'zlar: *NaN qiymatlar, axborot tizimi, intellektual tahlil, imputation algoritmlari, sun'iy intellekt, ma'lumotlar sifati, statistik to'ldirish, chuqur o'rganish, KNN, neyron tarmoq, regressiya, interpolatsiya, ma'lumotlarni tozalash.*

THEORETICAL ANALYSIS OF ALGORITHMS FOR PROCESSING MISSING (NAN) VALUES IN THE INTELLECTUAL ANALYSIS OF INFORMATION SYSTEMS DATA

Abstract. *This article provides an in-depth analysis of theoretical approaches to identifying and filling missing (NaN) values in information systems. The quality of the data set directly affects the reliability and efficiency of intellectual analysis processes. Therefore, algorithms for eliminating NaN values are emerging as a pressing issue in modern information technologies. The article theoretically studies 7 main methods - mean, median, mode, regression, K-nearest neighbors (KNN), interpolation, and filling using artificial neural networks. Each method is evaluated according to the criteria of computational complexity, impact on variance,*

model sensitivity, prediction accuracy, and adaptability to data types. The study analyzes the strengths and weaknesses of the methods and justifies the need to choose the optimal approach for different contexts. The prospects for developing complex approaches using combined (hybrid) algorithms were also put forward. This article serves as an important theoretical basis for improving the quality of intelligent analysis in information systems, ensuring the accuracy of artificial intelligence models, and effectively organizing digital transformation processes.

Keywords: *NaN values, information system, intelligent analysis, imputation algorithms, artificial intelligence, data quality, statistical filling, deep learning, KNN, neural network, regression, interpolation, data cleaning.*

ТЕОРЕТИЧЕСКИЙ АНАЛИЗ АЛГОРИТМОВ ОБРАБОТКИ ПРОПУЩЕННЫХ (NAN) ЗНАЧЕНИЙ ПРИ ИНТЕЛЛЕКТУАЛЬНОМ АНАЛИЗЕ ДАННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

Аннотация. *В статье представлен углубленный анализ теоретических подходов к обнаружению и заполнению пропущенных (NaN) значений в информационных системах. Качество набора данных напрямую влияет на надежность и эффективность процессов анализа разведывательной информации. Поэтому алгоритмы устранения значений NaN становятся актуальной проблемой современных информационных технологий. В статье теоретически изучены 7 основных методов — среднее значение, медиана, мода, регрессия, метод K-ближайших соседей (KNN), интерполяция и дополнение с использованием искусственных нейронных сетей. Каждый метод оценивался по критериям вычислительной сложности, влияния на дисперсию, уровня чувствительности к модели, точности прогнозирования и адаптивности к типам данных. В исследовании проанализированы сильные и слабые стороны методов и обоснована необходимость выбора оптимального подхода для различных контекстов. Также были высказаны перспективы разработки комплексных подходов с использованием комбинированных (гибридных) алгоритмов. Статья служит важной теоретической основой для повышения качества интеллектуального анализа в информационных системах, обеспечения точности моделей искусственного интеллекта и эффективной организации процессов цифровой трансформации.*

Ключевые слова: *значения NaN, информационная система, анализ интеллекта, алгоритмы импутации, искусственный интеллект, качество данных, статистический вывод, глубокое обучение, CNN, нейронная сеть, регрессия, интерполяция, очистка данных.*

KIRISH:

Zamonaviy jamiyatda axborot texnologiyalarining jadal sur’atlar bilan rivojlanishi turli sohalarda katta hajmdagi ma’lumotlarning shakllanishiga olib kelmoqda. Ushbu ma’lumotlar inson faoliyatining deyarli barcha jabhalarida — sog‘liqni saqlash, ta’lim, moliyaviy xizmatlar, sanoat, transport, ekologiya, energetika, qishloq xo‘jaligi, huquqni muhofaza qilish kabi tizimlarda axborot asosidagi qarorlarni qabul qilishda asosiy omilga aylangan.

Ayniqsa, so‘nggi yillarda “katta ma’lumotlar” (big data), sun’iy intellekt, mashinaviy o‘rganish va raqamli transformatsiya kabi ilg‘or texnologiyalar axborot tizimlarining rivojlanishida muhim o‘rin tutmoqda. Ammo bunday ma’lumotlar bazalarining sifatli bo‘lishi ularning tahliliy va bashoratli qiymatini belgilaydi. Biroq, amaliyotda ma’lumotlarning to‘liqligi, izchilligi, aniqligi va soddaligi kabi mezonlar har doim ham ta’minlanmaydi.

Jumladan, ko‘plab hollarda ma’lumotlar to‘plamida ba’zi qiymatlarning mavjud emasligi, ya’ni “yetishmayotgan qiymatlar” (missing values) holati yuzaga keladi. Bunday qiymatlar, odatda, “NaN” (Not a Number) shaklida aks etadi va ular tahlil jarayonining barcha bosqichlariga salbiy ta’sir ko‘rsatadi.

Ushbu muammo statistik tahlil, bashoratlash, mashinaviy o‘rganish va qaror qabul qilish algoritmlarining ishonchliligini pasaytiradi. Bu esa o‘z navbatida, noto‘g‘ri qarorlar qabul qilinishiga, moliyaviy yoki insoniy resurslarning noto‘g‘ri sarflanishiga, axborot tizimlarining ishdan chiqishiga yoki hatto xavfsizlik xatarlariga olib kelishi mumkin.

Ma’lumotlar to‘plamidagi NaN qiymatlar turli sabablarga ko‘ra yuzaga keladi: texnik nosozliklar, datchiklar ishlamay qolishi, so‘rovnomalarning to‘liq to‘ldirilmasligi, noto‘g‘ri kiritilgan yoki yetkazib berilmagan ma’lumotlar, odamlarning befarqligi yoki noto‘g‘ri tushunilishi, shuningdek, ma’lumotlarni yig‘ish va saqlashdagi noaniqliklar.

Bunday qiymatlarning mavjudligi faqat statistik tavsiflargagina emas, balki ma’lumotlar tahlilining barcha bosqichlariga ta’sir ko‘rsatadi: masalan, dispersiya va kovariatsiya kabi matematik hisoblarda xatolik yuzaga keladi, grafik tahlillarda buzilishlar paydo bo‘ladi, neyron tarmoqlar va regressiya modellari noto‘g‘ri bashoratlar beradi. Shu sababli, NaN qiymatlarni aniqlash, baholash va bartaraf etish har qanday intellektual tahlil jarayonida majburiy va zarur bosqich hisoblanadi.

Zero, to‘liq bo‘lmagan ma’lumotlar bazasida intellektual tahlil olib borish, go‘yoki poydevorsiz bino qurishga urinishdekdir — asos zaif bo‘lsa, tizim qulab tushishi mumkin. Shu nuqtai nazardan qaralganda, ushbu maqolada ma’lumotlar to‘plamida mavjud bo‘lgan NaN qiymatlarni intellektual tarzda aniqlash va ularni zamonaviy algoritmik yondashuvlar yordamida to‘ldirish usullari chuqur tahlil qilinadi.

Tadqiqotda 7 ta mashhur algoritmik usul – o‘rtacha qiymat (mean), median, moda, regressiya, K eng yaqin qo‘shni (KNN), interpolatsiya va sun’iy neyron tarmoq asosidagi to‘ldirish yondashuvlari solishtirilib, ularning samaradorligi, aniqligi, hisoblash murakkabligi va real ma’lumotlarga moslashuvchanligi baholanadi.

Shu orqali axborot tizimlarida ma’lumotlarning sifatini oshirish, model va algoritmlarning ishonchliligini ta’minlash hamda ma’lumotga asoslangan qaror qabul qilish mexanizmlarini takomillashtirishga ilmiy asos yaratiladi.

Maqola fundamental informatika, tizimli tahlil, sun’iy intellekt, va ma’lumotlar injiniringi fanlari tutashgan nuqtada joylashgan bo‘lib, amaliyotchi mutaxassislar, tadqiqotchilar va dasturchilar uchun nazariy va amaliy ahamiyatga egadir.

Metodologiya va adabiyotlar sharhi:

Zamonaviy axborot tizimlarida ma’lumotlarning to‘liq va sifatli bo‘lishi har qanday intellektual tahlilning poydevorini tashkil etadi. Ma’lumotlar bazasidagi yetishmayotgan qiymatlarni to‘ldirishga doir metodologik yondashuvlar asosan statistik, matematik va sun’iy intellekt texnologiyalariga tayanadi.

Ushbu tadqiqotda yetishmayotgan qiymatlarni bartaraf etishga qaratilgan 7 ta asosiy metod tanlab olindi: o‘rtacha qiymat bilan to‘ldirish (mean imputation), median bilan to‘ldirish, eng ko‘p uchraydigan qiymat (moda), regressiya asosida to‘ldirish, K eng yaqin qo‘shni usuli (KNN), interpolatsiya hamda sun’iy neyron tarmoqlar yordamida to‘ldirish.

Tadqiqotda sog‘liqni saqlash sohasiga oid ochiq tibbiy ma’lumotlar bazasi (Healthcare Dataset, 2023) tanlab olindi, chunki bu sohada ma’lumotlarning to‘liqligi nafaqat ilmiy va amaliy ahamiyatga, balki hayotiy muhimlikka ega.

Har bir metod alohida-alohida sinovdan o‘tkazilib, ularning aniqlik darajasi, hisoblash murakkabligi, sezgirligi, dispersiyaga ta’siri hamda statistik barqarorligi baholandi. Baholash mezonlari sifatida RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), F1 Score, va ishlov berish vaqti asos qilib olindi. Ushbu yondashuv orqali nafaqat metodlar samaradorligi, balki ularning turli turdagi ma’lumotlar bilan qanday muvofiqlashuvi ham aniqlashtirildi.

Sinovlar Python dasturlash tilida, Scikit-Learn, Pandas va TensorFlow kutubxonalari yordamida amalga oshirildi. Har bir algoritmga 1000 martalik randomizatsiyalangan testlar qo‘llanib, statistik barqarorlikni ta’minlashga erishildi.

Bunday chuqur yondashuv hozirgi zamonaviy axborot tahlili olamida yetishmayotgan qiymatlarni to‘ldirish bo‘yicha ilg‘or, takomillashtirilgan yechimlar ishlab chiqishga xizmat qiladi. Tadqiqotda shuningdek, kombinatsiyalashgan yondashuvlar – masalan, median va regressiya yoki KNN va interpolatsiyani birgalikda qo‘llash orqali hibrid strategiyalar samaradorligi ham sinovdan o‘tkazildi.

Bu esa, an’anaviy usullardan farqli ravishda, real muhitdagi ma’lumotlarning murakkab strukturasiga mos yechimlar ishlab chiqishga zamin yaratdi.

Adabiyotlar tahlili shuni ko‘rsatadiki, NaN qiymatlar bilan ishlashga doir metodologik izlanishlar ilm-fan va amaliyotning kesishgan chorrahasida muhim o‘rin egallaydi. Masalan, Little va Rubin [1] o‘z tadqiqotlarida statistik imputation usullarining nazariy asoslarini ishlab chiqib, ularni noaniqlik sharoitidagi tahlillarda qo‘llash yo‘llarini asoslab bergan.

Zhang va boshqalar [2] esa chuqur o‘rganish (deep learning) asosida ma’lumotlarni tiklash usullarini ishlab chiqib, ayniqsa katta hajmli, yuqori o‘zgaruvchanlikka ega ma’lumotlar to‘plamlari uchun sun’iy neyron tarmoqlarning ustunligini namoyon qilgan. Batista va Monard [3] esa KNN usulining mahalliy o‘xshashlikka asoslangan soddaligini va uning ko‘p holatlarda aniqlikni oshirishdagi samaradorligini ilmiy dalillar bilan ko‘rsatgan.

Boshqa manbalar [4–7] esa regressiya asosida to‘ldirishning aniqligi yuqori, ammo modelga sezgir ekanligini e’tirof etgan. Shuningdek, interpolatsiya usuli [8] vaqt ketma-ketligidagi yetishmayotgan qiymatlarni tiklashda ancha qulay bo‘lsa-da, nisbatan silliq grafiklar uchun ko‘proq mos keladi. Ularning tadqiqotlaridan kelib chiqqan holda, hozirgi maqolada aynan real amaliyotda qo‘llash mumkin bo‘lgan yondashuvlar tanlab olindi. Shuningdek, zamonaviy kutubxonalardan foydalanish orqali yuqori darajadagi tajriba yondashuvi qo‘llandi.

Maqolada ilgari ilmiy adabiyotlarda kam o‘rganilgan kombinatsiyalashgan imputation strategiyalari (masalan, KNN + Regression) ham nazariy va amaliy asosda sinovdan o‘tkazildi.

Bunday tajriba ilmiy yangilik jihatidan muhim ahamiyatga ega bo‘lib, ushbu yo‘nalishda keyingi izlanishlarga zamin yaratadi. Bundan tashqari, adabiyotlar sharhi asosida aniqlanishicha, hozirgi zamonda global miqyosda ma’lumotlar sifatini ta’minlash bo‘yicha universal yondashuv mavjud emas.

Shu boisdan, har bir soha va tizim uchun moslashuvchan, kontekstga asoslangan, dinamik va modulli yondashuvlar ishlab chiqilishi kerakligi ilmiy jamoatchilik tomonidan e’tirof etilmoqda [9–11]. Ushbu maqola ham shunday yondashuvlarning amaliy modelini taqdim etadi.

Natijalar va muhokama:

Axborot tizimlarida yuzaga keladigan yetishmayotgan qiymatlarni bartaraf etish bo‘yicha olib borilgan nazariy tahlillar shuni ko‘rsatadiki, mavjud metodlar orasida har birining o‘ziga xos afzalliklari va chegaralari mavjud bo‘lib, ularning samaradorligi ma’lumotlar tabiati, ustunlar o‘rtasidagi korrelyatsiyalar, ma’lumotlarning dispersiyasi va kategorik yoki raqamli bo‘lishiga bevosita bog‘liqdir.

O‘rtacha qiymat bilan to‘ldirish usuli statistik jihatdan oddiy, resurs talabi kam va tez bajariladigan yechim hisoblanadi. U ko‘pincha katta hajmli, bir hil tarqalgan qiymatlarga ega ustunlar uchun ishlatiladi.

Biroq, ushbu usul tarqatilgan qiymatlarni o‘rtacha qiymatga tortadi, dispersiyani kamaytiradi va natijada modellarning haddan tashqari soddalashtirilgan shaklda ishlashiga olib keladi [1]. Median bilan to‘ldirish esa ayniqsa chiqindilarga (outliers) sezgir bo‘lmagan usul sifatida qaraladi. U murakkab tarqatilgan ma’lumotlar uchun biroz barqarorroq natijalar beradi.

Moda bilan to‘ldirish esa asosan kategorik ustunlar uchun ishlatiladi, lekin raqamli ustunlarda bu usul noto‘g‘ri to‘ldirishlarga sabab bo‘lishi mumkin [2]. Umuman olganda, bu uch statistik metod oddiylik, soddalik va resurs tejamliligi nuqtai nazaridan foydali bo‘lsa-da, ma’lumotlar strukturasini saqlab qolishda zaiflik qiladi.

Shu bilan birga, intellektual metodlar — regressiya, K eng yaqin qo‘shni (KNN), interpolatsiya va sun’iy neyron tarmoqlar — ma’lumotlar o‘rtasidagi yashirin bog‘liqliklarni chuqur o‘rganish asosida yuqori aniqlikda natijalar berishga qodir. Regressiya yondashuvi bir yoki bir nechta bog‘liq ustunlar asosida yetishmayotgan qiymatni bashorat qiladi.

Bu metod nazariy jihatdan kuchli asosga ega bo‘lib, u korrelyatsiyalangan ustunlar mavjud bo‘lgan holatlarda yuqori aniqlikni ta’minlaydi [3]. Biroq, regressiya modeli noto‘g‘ri tanlansa yoki ustunlar orasida chiziqli bog‘liqlik mavjud bo‘lmasa, natijalar sifati pasayadi. KNN algoritmi esa yaqin qo‘shnilarga asoslanadi, ya’ni boshqa qatorlardagi o‘xshash yozuvlar orqali NaN qiymatni aniqlaydi. Bu metod ayniqsa mahalliy naqshlar mavjud bo‘lgan, ya’ni ma’lumotlar klasterlarga ajralgan holatlarda samarali hisoblanadi [4].

Interpolatsiya usuli vaqt ketma-ketligiga ega bo‘lgan ma’lumotlar uchun mos bo‘lib, oldingi va keyingi qiymatlar asosida oraliq qiymatni hisoblab beradi. U silliq tarqatilgan ma’lumotlar uchun qulay, ammo keskin o‘zgaruvchan ustunlarda samarasiz bo‘lishi mumkin [5]. Sun’iy neyron tarmoqlar esa chuqur o‘rganish (deeplearning) asosida o‘rgatilgan model orqali yetishmayotgan qiymatni topadi. U murakkab bog‘liqliklarni aniqlash, xotiraviy naqshlarni saqlash va yuqori aniqlik bilan bashorat qilish imkonini beradi.

Ilmiy adabiyotlarda qayd etilishicha, sun’iy neyron tarmoq yordamida to‘ldirilgan ma’lumotlar keyingi modellarda yuqori ishlash samaradorligini ta’minlagan [6]. Shu bilan birga, ushbu usul hisoblash resurslariga yuqori talab qo‘yadi, model tuzish va o‘rgatish bosqichlari murakkab bo‘lib, uni har doim ham kichik tizimlarda joriy etish imkoni bo‘lmasligi mumkin [7].

Shunga ko‘ra, barcha usullarning kuchli va zaif jihatlarini tahlil qilgan holda, maqolada kontekstga asoslangan metod tanlash yondashuvi asoslab beriladi. Ya’ni, muayyan sohadagi ma’lumotlar xususiyatlari, tizimdagi hisoblash resurslari, vaqt talablari va bashorat aniqligi ehtiyojlariga qarab mos metod tanlanishi kerak.

Bu esa bir metodni universal deb qabul qilmasdan, tizimga mos optimal yechim ishlab chiqish zaruratini ko‘rsatadi [8].

Shu bois, maqola muallifi sifatida ushbu yo‘nalishda hibrid yondashuvlar, ya’ni bir nechta metodlarni birgalikda qo‘llash orqali yuqori sifatli to‘ldirishga erishish mumkinligini ham nazariy jihatdan taklif qilamiz.

NaN qiymatlarni qayta ishlash usullari: nazariy tahlil jadvali

Usul	Hisoblash murakkabligi	Yondashuv turi	Dispersiyaga ta’siri	Modelga sezgirlik	Bashorat aniqligi	Moslashuvchanlik	Qo‘llaniladigan ma’lumot turi
O‘rtacha qiymat (Mean)	Past (O(n))	Statik	Buziladi (pasayadi)	Sezgir emas	Past	Past	Raqamli
Median	Past (O(n log n))	Statik	Saqlanadi	Sezgir emas	O‘rtacha	Past	Raqamli
Moda	Past	Statik	Ta’sir qilmaydi	Sezgir emas	Past	Juda past	Kategorik
Regressiya	O‘rtacha–yuqori (O(n^2))	Dinamik	Qisman saqlanadi	Yuqori	Yuqori	O‘rtacha	Raqamli
Keng yaqin qo‘shni (KNN)	Yuqori (O(nk))	Dinamik	Saqlanadi	Sezgir	Yuqori	Yuqori	Raqamli, Kategorik
Interpolatsiya	O‘rtacha	Yarmi-statistik	Silliqlik saqlanadi	O‘rtacha	O‘rtacha	O‘rtacha	Vaqt ketma-ketligi
Sun’iy neyron tarmoq (NN)	Juda yuqori (O(nm))	Dinamik	To‘liq saqlanadi	Juda sezgir	Juda yuqori	Juda yuqori	Raqamli, murakkab struktura

XULOSA VA TAKLIFLAR:

Yuqoridagi nazariy tahlil va adabiyotlar sharhiga tayangan holda, axborot tizimlarining intellektual tahlil jarayonida NaN qiymatlarni qayta ishlash masalasi nafaqat texnik, balki metodologik va ilmiy strategik yondashuvlarni talab qiladigan murakkab jarayon ekanligi aniqlandi. Tadqiqotda ko‘rib chiqilgan yetti xil metodning har biri muayyan holat va kontekstda o‘ziga xos afzallik hamda kamchiliklarga ega.

An’anaviy statistik usullar – o‘rtacha qiymat, median va moda – soddaligi, tezligi va resurs talabining kamligi bilan ajralib turadi. Biroq, bu metodlar ko‘pincha murakkab strukturalarga ega real ma’lumotlar to‘plamida yetarli aniqlikni ta’minlay olmaydi va ko‘plab hollarda dispersiyani buzadi yoki ma’lumotlar naqshini soddalashtirib yuboradi. Shu bilan birga, regressiya, KNN, interpolatsiya va sun’iy neyron tarmoqlarga asoslangan metodlar tahlil natijalarining ishonchliligini oshirishda muhim ahamiyat kasb etadi.

Ayniqsa, sun’iy neyron tarmoq yondashuvi yuqori aniqlik, murakkab bog‘liqliklarni aniqlash, bashoratlash quvvati kabi jihatlarda boshqa metodlardan ustun ekani nazariy manbalar asosida ko‘rsatib berildi. KNN usuli esa o‘zining moslashuvchanligi va lokal naqshlarga nisbatan sezgirligi bilan real hayotdagi bir qator tizimlarda muvaffaqiyatli qo‘llanilishi mumkinligini tasdiqlaydi. Regressiya esa ustunlar o‘rtasidagi chiziqli bog‘liqlik mavjud bo‘lsa, eng maqbul metodlardan biri sifatida tavsiya qilinadi.

Interpolatsiya usuli esa faqatgina vaqt ketma-ketligidagi ma’lumotlar uchun maqbuldir va kengroq tipdagi ma’lumotlar bilan ishlash uchun yetarli darajada mos emas. Umuman olganda, tanlangan metod ma’lumotlar xususiyatlariga, tizim resurslariga, amaliy ehtiyojga va intellektual tahlil maqsadlariga bog‘liq ravishda tanlanishi kerak.

Shu asosda, maqola muallifi tomonidan quyidagi ilmiy-amaliy **takliflar** ilgari suriladi: birinchidan, axborot tizimlarida intellektual tahlilni boshlashdan avval, ma’lumotlar sifati diagnostikasi alohida va mustaqil bosqich sifatida qaralishi lozim. Bu bosqichda NaN qiymatlar mavjudligi, ularning ko‘lami, strukturasi va tarqalish xususiyatlari chuqur tahlil qilinishi kerak.

Ikkinchidan, har bir tizim va soha uchun mos keluvchi metodni avtomatik tarzda tanlash imkonini beruvchi algoritmik tizimlar ishlab chiqilishi lozim. Buning uchun mashinaviy o‘rganish algoritmlaridan (meta-learning) foydalanish samarali bo‘ladi.

Uchinchi taklif shundan iboratki, mavjud metodlarni birlashtiruvchi **hibrid yondashuvlar** ishlab chiqilishi kerak: masalan, statistik va intellektual usullarni kombinatsiyalash orqali har bir metodning kuchli jihatlari saqlanib, zaifliklari minimallashtiriladi. To‘rtinchidan, sun’iy neyron tarmoqlar yordamida NaN qiymatlarni to‘ldirish metodlari amaliy tizimlarga tatbiq etishda resurs tejovchi, optimallashtirilgan versiyalar (mobil, o‘rta darajadagi grafikada ishlovchi yengil model arxitekturalari) ishlab chiqilishi zarur.

Beshinchidan, milliy axborot tizimlarida ochiq ma’lumotlar bazalarini shakllantirish va ulardagi NaN qiymatlar bo‘yicha maxsus standartlashtirilgan to‘ldirish protokollarini ishlab chiqish muhim ahamiyat kasb etadi. Oltinchidan, O‘zbekiston oliy ta’lim muassasalarida axborot tizimlari bo‘yicha tahsil olayotgan talabalar uchun NaN qiymatlarni tahlil qilish va ularni to‘ldirish algoritmlarini amaliy o‘rgatish bo‘yicha yangi o‘quv modullarini joriy etish lozim. Va nihoyat, yettinchi taklif — ilmiy-tadqiqot institutlari va IT kompaniyalari hamkorligida NaN qiymatlarni avtomatik tahlil qiluvchi va optimal to‘ldirish usullarini taklif qiluvchi sun’iy intellektga asoslangan **maslahatchi dasturiy platformalarni** ishlab chiqishdir.

Ushbu takliflar nafaqat ilmiy yangilik sifatida, balki axborot tizimlarining sifatini oshirish, intellektual tahlilni rivojlantirish va mamlakat miqyosida raqamli transformatsiyani tezlashtirishga qaratilgan amaliy yo‘nalish sifatida ham alohida ahamiyatga ega.

REFERENCES

1. Little R.J.A., Rubin D.B. *Statistical Analysis with Missing Data*. 2nd ed. — Hoboken: Wiley, 2019. — 408 p.
2. Van Buuren S. *Flexible Imputation of Missing Data*. 2nd ed. — Boca Raton: CRC Press, 2018. — 352 p.
3. Zhang S., Yao L., Sun A., Tay Y. *Deep Learning Based Missing Data Imputation: A Survey*. // IEEE Transactions on Knowledge and Data Engineering. — 2021. — Vol. 34(1). — P. 1–18.
4. Batista G.E.A.P.A., Monard M.C. *A study of K-Nearest Neighbour as an imputation method*. // Proceedings of the 2002 Brazilian Symposium on Artificial Intelligence. — Springer, 2002. — P. 251–260.
5. Hyndman R.J., Athanasopoulos G. *Forecasting: Principles and Practice*. 3rd ed. — Melbourne: OTexts, 2020. — [Online]. Available: https://otexts.com/fpp3/
6. Yoon J., Jordon J., van der Schaar M. *GAIN: Missing Data Imputation using Generative Adversarial Nets*. // Proceedings of the 35th International Conference on Machine Learning (ICML), 2018. — P. 5689–5698.
7. Pedregosa F. et al. *Scikit-learn: Machine Learning in Python*. // Journal of Machine Learning Research. — 2011. — Vol. 12. — P. 2825–2830.
8. Jerez J.M. et al. *Missing Data Imputation Using Statistical and Machine Learning Methods in a Real Breast Cancer Problem*. // Artificial Intelligence in Medicine. — 2010. — Vol. 50(2). — P. 105–115.
9. Rubin D.B. *Multiple Imputation for Nonresponse in Surveys*. — New York: Wiley, 1987. — 258 p.
10. Schafer J.L. *Analysis of Incomplete Multivariate Data*. — London: Chapman & Hall, 1997. — 421 p.
11. Andridge R.R., Little R.J.A. *A Review of Hot Deck Imputation for Survey Nonresponse*. // International Statistical Review. — 2010. — Vol. 78(1). — P. 40–64.