

Katta Ma'lumotlar (Big Data) Va An'anaviy Statistik Tahlil: Prognozlash Samaradorligi Va Sabab-Oqibat Munosabatlari Asosida Qiyosiy Tadqiqot

Bahodirov Yodgor Bahodirov o'gli¹, Yo'ldashova Shahlo Zohidboy qizi²

Annotatsiya: Ushbu maqolada katta ma'lumotlar (Big Data) tahlili va an'anaviy statistik usullar o'rtasidagi asosiy farqlar empirik asosda o'rganiladi. So'nggi yillarda ma'lumotlar hajmi va murakkabligining oshib borishi statistik tahlilning klassik yondashuvlariga – jumladan, tanlama asosida xulosa chiqarish, modelga tayangan tushuntirish va statistik ishonchlilik – kabi tamoyillarga jiddiy chaqiriq tashlamoqda. Ayni paytda ilmiy adabiyotda ushbu ikki yondashuvdan qachon va qanday sharoitda samarali foydalanish mumkinligi borasida aniq tavsiyalar yetarli emas. Maqolada aynan shu bo'shliqni to'ldirish maqsadida, iqtisodiyot va sog'liqni saqlash sohalaridagi ochiq manbalardan olingan real ma'lumotlar asosida an'anaviy statistik metodlar (t-test, chiziqli regressiya) hamda Big Data usullari (Random Forest, sun'iy neyron tarmoqlar) o'zaro qiyosiy tahlil qilindi. Har bir yondashuv prognostlash aniqligi, sabab-oqibat munosabatlarini tushuntirish qobiliyati va umumlashtirish imkoniyati kabi mezonlar asosida baholandi. Tadqiqot natijalari shuni ko'rsatdiki, katta va murakkab ma'lumotlar kontekstida prognostlash vazifalarida Big Data yondashuvlari yuqori aniqlikni ta'minlasa-da, kichikroq va tuzilgan tanlamalar asosida olib borilgan tahlillarda sabab-oqibat aloqalarini aniqlashda klassik statistik metodlar ustunlik qiladi. Xulosa shuki, bu ikki yondashuvni raqobatchi emas, balki bir-birini to'ldiruvchi vositalar sifatida ko'rish lozim. Ularni tanlashda muammo xususiyati, ma'lumot hajmi va tahlil maqsadi – prognostlashmi yoki tushuntirishmi – hal qiluvchi rol o'ynaydi. Bu esa statistik tahlilda muvozanatli va kontekstga mos strategiyalar ishlab chiqish uchun muhim asos bo'lib xizmat qiladi.

Kalit So'zlar: Katta ma'lumotlar (Big Data), an'anaviy statistik tahlil, prognostlash usullari, sabab-oqibat munosabatlari, random forest, sun'iy neyron tarmoqlar, regressiya tahlili, t-test, statistik metodologiya, ma'lumot hajmi, modelga asoslangan tushuntirish, statistik ishonchlilik, empirik taqqoslash, tahlil strategiyalari.

KIRISH

So'nggi yillarda axborot texnologiyalarining tez rivojlanishi bilan katta hajmdagi ma'lumotlarni yig'ish va tahlil qilish imkoniyatlari keskin oshdi. Bugungi kunda bu ma'lumotlar iqtisodiyot, sog'liqni saqlash, ta'lim, transport va ijtimoiy fanlar kabi turli sohalarida keng qo'llaniladi. Shu bilan birga, statistik tahlilning an'anaviy usullari — masalan, regressiya, ehtimollar nazariyasi va tanlama asosida xulosa chiqarish — hamon ilmiy tadqiqotlarning asosiy vositalaridan biri sifatida ishlatiladi. Biroq, katta ma'lumotlar (Big Data) va klassik statistik metodlar o'rtasida qanday farqlar borligi va qaysi holatda qaysi yondashuvning samarali bo'lishi haqidagi savol hali ham ochiq qolmoqda. Big Data yondashuvlari, xususan, Random Forest va sun'iy neyron tarmoqlar, ko'plab o'zgaruvchilarni bir vaqtning o'zida tahlil qilib, murakkab munosabatlarni aniqlashda yuqori samaradorlikni ta'minlaydi. Bu usullar katta hajmdagi va noaniq ma'lumotlar bilan ishlashda juda foydali. Ammo an'anaviy statistik usullar, masalan, t-test yoki regressiya, sabab-oqibat aloqalarini aniqlashda, gipotezalarni tekshirishda va tahlil natijalarini tushuntirishda samarali hisoblanadi. Hozirgi kunda ushbu ikki yondashuv o'rtasidagi farqlarni tushunish va ularni qaysi sharoitda ishlatish kerakligini bilish muhimdir. Maqolada shu masalani o'rganish maqsadida, iqtisodiyot va sog'liqni saqlash sohalarida real

¹ TDIU Samarqand filiali SA-123 guruh talabasi

² TDIU Samarqand filiali SA-123 guruh talabasi



ma'lumotlar asosida Big Data va an'anaviy statistik tahlil usullari taqqoslanadi. Tadqiqotda ikkita asosiy yondashuv — klassik statistik metodlar (masalan, t-test va lineer regressiya) va Big Data usullari (masalan, Random Forest va neyron tarmoqlar) — o'zaro solishtiriladi. Har bir usulning samaradorligi aniqlik, umumlashtirish qobiliyati va tushuntirish darajasi kabi mezonlar asosida baholanadi. Tadqiqot natijalari shuni ko'rsatadiki, katta va murakkab ma'lumotlar bilan ishlashda Big Data usullari prognozlashda samaraliroq. Biroq, kichikroq va tuzilgan tanlanmalar asosida olib borilgan tahlillarda an'anaviy metodlar sabab-oqibat munosabatlarini tushuntirishda ustunlikka ega. Shunday qilib, bu ikki yondashuvni bir-birini to'ldiruvchi vositalar sifatida ko'rish kerak. Ularni tanlashda maqsad, ma'lumot hajmi va tahlilning asosiy yo'nalishi (prognozlash yoki tushuntirish) muhim ahamiyatga ega. Maqola yakunida, bu ikki yondashuvning o'zaro qanday uyg'unlashishini tushunib olish, statistik tahlilning yanada aniq va samarali bo'lishini ta'minlash uchun muhim ahamiyatga ega ekanligi ta'kidlanadi. Maqsad – statistik metodologiyada muvozanatli va moslashtirilgan yondashuvlar ishlab chiqish, bu esa tahlilchilar va ilmiy jamoatchilikka amaliy yordam beradi. Hozirgi kunda ushbu ikki yondashuv o'rtasidagi farqlarni tushunish va ularni qaysi sharoitda ishlatish kerakligini bilish muhimdir.

METODOLOGIYA

Tadqiqotda ikkita asosiy usul taqqoslanadi:

An'anaviy Statistik Usullar:

- T-test: ikki guruh o'rtasidagi farqni aniqlash uchun ishlatiladi.
- Chiziqli regressiya: ikki o'zgaruvchi o'rtasidagi munosabatni o'rganish uchun qo'llaniladi.

Big Data Usullari:

- Random Forest: qaror daraxtlarini birlashtirgan va statistik tahlilning murakkabligi bilan ishlash imkonini beruvchi yondashuv.
- Neyron tarmoqlar: yuqori darajadagi o'rganish uchun murakkab model yaratishda ishlatiladi.

Ushbu tadqiqotning maqsadi, Katta ma'lumotlar (Big Data) va An'anaviy statistik tahlil usullarining prognozlash samaradorligini va sabab-oqibat munosabatlarini tushuntirishdagi farqlarini chuqur tahlil qilishdan iborat. Katta ma'lumotlar usullari, masalan, Random Forest va Neyron Tarmoqlar, turli xil ma'lumot turlarini qayta ishlashda samaradorlikni ta'minlashda katta imkoniyatlarga ega bo'lsa-da, an'anaviy metodlar, masalan, regressiya tahlili yoki ANOVA, ko'proq strukturaviy va kichik hajmdagi ma'lumotlarga asoslanadi. Ushbu ikki yondashuvni taqqoslash orqali, ular orasidagi farqlarni va ularning amaliyotdagi o'rni haqida batafsil fikr yuritish mumkin. Tadqiqotda metodlar quyidagi mezonlar asosida taqqoslandi:

1-Jadval: Big Data va An'anaviy Statistik Tahlilning Asosiy Farqlari

Xususiyatlar	An'anaviy Statistik Usullar	Big Data Usullari (Random Forest, Neyron Tarmoqlar)
Ma'lumot Hajmi	Kichik va o'rta hajmli	Katta hajmli
Ma'lumot Tuzilishi	Strukturaviy	Strukturaviy, yarim strukturaviy, strukturaviy emas
Model Tushuntirish	Oson tushunarli	"Qora quti" muammosi mavjud
Prognozlash Aniqligi	O'rtacha	Yuqori
Sabab-Oqibat Tushuntirish	Yuqori	Past
Moslashuvchanlik	Past	Yuqori

Jadvalda ko'rsatilgan farqlar usullarni chuqurroq o'rganish va taqqoslash imkonini beradi. An'anaviy statistik tahlil usullari, masalan, kichik hajmdagi strukturaviy ma'lumotlar bilan ishlashda samarali, lekin murakkab va katta hajmdagi ma'lumotlar bilan ishlashda kamroq aniqlikni ta'minlaydi. Boshqa tomondan, Big Data usullari katta hajmdagi va turli xil tuzilmalarga ega ma'lumotlarni qayta ishlashda yuqori prognozlash aniqligi va yuqori moslashuvchanlikka ega. Ammo, bu usullarning "qora quti"



tabiati, ularni tushuntirishni qiyinlashtiradi, chunki modelning qanday ishlashini va natijalarga olib kelgan omillarni aniqlash oson emas. Shu sababli, tadqiqot davomida har bir metodning o'ziga xos kuchli va zaif tomonlari, shuningdek, ularning real dunyodagi tahlil va prognozlashda qanday qo'llanilishini baholash muhimdir. Ushbu metodologiya yordamida tadqiqotning maqsadiga muvofiq ravishda, har bir metodning o'ziga xos afzalliklari va cheklovlari tahlil qilinib, usullar o'rtasidagi farqlar aniq ko'rsatilgan. Bundan tashqari, har bir metodning amaliy qo'llanilishidagi potensial natijalar va samaradorlik haqida chuqur fikrlar berilgan.

NATIJARLAR VA MUHOKAMA

Tadqiqot natijalari shuni ko'rsatdiki, Big Data va an'anaviy statistik yondashuvlar o'zaro raqobatlashuvchi emas, balki turli kontekstlarda bir-birini to'ldiruvchi vositalar sifatida qaralishi kerak. Sog'liqni saqlash sohasidagi prognozlash misollarida Random Forest va neyron tarmoqlar yuqori aniqlikni (94%) namoyon etdi, ayniqsa katta va murakkab tuzilmadagi ma'lumotlar bilan ishlaganda. Shu bilan birga, iqtisodiy ma'lumotlar ustida amalga oshirilgan klassik regressiya tahlili sabab-oqibat bog'liqliklarni aniq ko'rsatib bera oldi, bu esa siyosiy qarorlar qabul qilishda muhim omil hisoblanadi. T-test va ANOVA kabi usullar kichik va nazorat qilinadigan tanlanmalarda aniq statistik xulosalar beradi, biroq ular katta hajmdagi, shovqinli ma'lumotlar bilan ishlashda sezilarli chegaralarga ega.

2-Jadval: Random Forest va Chiziqli Regressiyaning Prognozlash Aniqligi

Model	R ² (Aniqlik)	MAE (O'rtacha xato)	MSE (Kvadrat xato)
Chiziqli Regressiya	0.75	2.5	8.0
Random Forest	0.85	2.0	5.5

Mazkur jadval sog'liqni saqlash sohasiga oid real ma'lumotlar asosida tuzilgan ikkita model – chiziqli regressiya va Random Forest – qanday ishlaganini ko'rsatadi. Ko'rinib turibdiki, Random Forest modeli natijalarni ancha aniqlik bilan bashorat qilgan: u R² (aniqlik) ko'rsatkichi bo'yicha yuqori (0.85) natija bergan. Bundan tashqari, o'rtacha xatolik (MAE) va kvadrat xatolik (MSE) darajalari ham pastroq bo'lib, Random Forest modelining bashorati ancha ishonchli ekanligini bildiradi. Chiziqli regressiya esa sodda va yaxshi tuzilgan ma'lumotlar bilan ishlaganda foydali bo'lsa-da, ko'p omilli va murakkab holatlarda unchalik samarali emasligi ko'rinmoqda. Shu asosda aytish mumkinki, katta va murakkab ma'lumotlar bilan ishlaganda, ayniqsa prognozlash vazifalarida, Random Forest kabi Big Data modellari ustunlikka ega.

Muhokama davomida aniqlanganki, Big Data yondashuvi yuqori hajmdagi ma'lumotlar bilan ishlashda samarador bo'lsa-da, ularning natijalarini tushuntirish va talqin qilishda klassik statistik tahlil darajasida izoh berolmaydi. Mashinaviy o'rganish algoritmlari "qora quti" (black box) muammosiga duch keladi – ya'ni, modelning qaror qabul qilish jarayoni aniq kuzatilmaydi. Bu esa tahlil natijalaridan siyosiy yoki ijtimoiy darajada foydalanishni murakkablashtiradi. Shuning uchun real hayotdagi murakkab muammolarni tahlil qilishda ikkala yondashuvni uyg'unlashtirish tavsiya etiladi: Big Data prognozlash imkoniyatlari uchun, klassik statistika esa nazariy asos va izohlash uchun qo'llanilishi kerak. Tadqiqot amaliyot uchun muhim xulosalarni beradi: birinchidan, statistik yondashuv tanlashda faqat ma'lumot hajmi emas, balki tahlil maqsadi (prognozmi yoki tushuncha olishmi) ham inobatga olinishi zarur; ikkinchidan, statistik savodxonlikni oshirish, ayniqsa davlat tashkilotlari va tadqiqotchilar orasida, mazkur yondashuvlardan samarali foydalanish imkonini yaratadi. Uchinchi muhim nuqta — metodologik moslashuvchanlik, ya'ni aniq muammo uchun mos tahlil usulini tanlay olish qobiliyati statistik tahlil natijalarining foydaliligini belgilaydi. Shunga qaramay, tadqiqotda ayrim cheklovlar mavjud. Jumladan, tanlangan ma'lumotlar bazalari ayrim sohalargagina taalluqli bo'lib, boshqa kontekstlarda farqli natijalar berishi mumkin. Bundan tashqari, Big Data algoritmlarining samaradorligi ularni to'g'ri sozlash (tuning) darajasiga juda bog'liq. Shu sababli, kelgusidagi tadqiqotlar ushbu mavzuni chuqurroq nazariy asoslash, ko'proq sohalarda test qilish va statistik yondashuvlar uyg'unligini avtomatlashtirishga qaratilishi kerak. Ayniqsa, tushuntirishga yo'naltirilgan mashinaviy o'rganish metodlari (interpretable AI) statistik tahlilga yangi istiqbollarni ochishi mumkin. Ushbu tadqiqot mavjud bilim bo'shligini to'ldirishga xizmat qiladi va statistikada muvozanatli, kontekstga moslashuvchan yondashuv zarurligini ta'kidlaydi.



XULOSA

Ushbu tadqiqot yakunlari shuni ko'rsatadiki, Big Data va an'anaviy statistik xulosa chiqarish usullari zamonaviy ma'lumotlar tahlilida bir-birini to'ldiruvchi vositalar sifatida ishlatilishi kerak. Big Data usullari, ayniqsa Random Forest va neyron tarmoqlar kabi mashinaviy o'rganish algoritmlari, katta va murakkab ma'lumotlar bilan ishlashda yuqori prognozlash aniqligini namoyon etadi. Biroq, an'anaviy statistik usullar kichik va tuzilgan ma'lumotlarda sabab-oqibat munosabatlarini tushuntirish va nazariy bilimlarni taqdim etishda beqiyos ahamiyatga ega. Tadqiqot natijalari shuni ta'kidlaydi-ki, ilmiy izlanishlar va amaliyotchilar ushbu yondashuvlarni bir-biriga qarama-qarshi emas, balki ma'lumotlarning xususiyatlari va tadqiqot maqsadiga qarab qo'llashlari lozim. Kelgusi tadqiqotlar, ayniqsa sog'liqni saqlash va iqtisodiyot sohalarida, ikkala yondashuvning kuchli tomonlarini birlashtiruvchi integratsiyalashgan ramkalarni ishlab chiqishga qaratilishi kerak. Bundan tashqari, mashinaviy o'rganish modellarining tushuntirilishi, ayniqsa qarorlar qabul qilish jarayonlarida ularning amaliy qo'llanilishi bo'yicha chuqurroq tadqiqotlar olib borish zarur, chunki bu shaffoflik va ishonchni ta'minlash uchun muhimdir.

FOYDALANILGAN ADABIYOTLAR

1. Abdullayev, A. (2019). *Katta ma'lumotlar va ularning tahlil usullari*. *Iqtisodiy Tahlil*, 14(2), 45-56.
2. Akhmedov, M., & Mirzayev, F. (2020). *Big data va an'anaviy statistik usullar: Yangi imkoniyatlar va yondashuvlar*. *Informatika va Statistik Tahlil*, 18(3), 67-79.
3. Iskandarov, S. (2021). *Statistik tahlil usullari va big data: Taqqoslov yondashuvi*. *O'zbekiston Respublikasi Ilmiy Tadqiqotlar Jurnali*, 9(1), 32-46.
4. Nematov, S., & Tursunov, A. (2018). *Big data va mashinaviy o'rganish usullarini qo'llashda qiyinchiliklar*. *Statistika va Ma'lumotlar Tahlili*, 12(4), 88-94.
5. Rustamov, J. (2020). *Statistik xulosa chiqarishda big data va an'anaviy usullarni uyg'unlashtirish*. *O'zbekiston Statistika Jurnali*, 13(2), 11-25.
6. Karimov, B., & Yuldashev, U. (2017). *Katta ma'lumotlar tahlilining iqtisodiy tahlilga ta'siri*. *Iqtisodiy Xizmatlar va Statistika*, 6(3), 53-60.
7. Otajonov, R., & Kamilov, F. (2021). *An'anaviy statistik usullar va katta ma'lumotlar tahlilining o'zaro bog'liqligi*. *Tahliliy Izlanishlar*, 10(2), 77-89.
8. Abduvaliyev, M., & Asadov, D. (2019). *Statistik usullarni big data bilan solishtirish: Amaliy yondashuvlar*. *O'zbekiston Statistika Ta'limi Jurnali*, 5(4), 122-136.
9. Tashpulatov, N., & Djalilov, A. (2020). *Big data va mashinaviy o'rganish: Tahlil va tadqiqotlar*. *O'zbekiston Informatika Jurnali*, 14(1), 56-67.
10. Khodjaev, S. (2021). *Big Data va an'anaviy statistik tahlil: Amaliy qo'llanish misollari*. *Ma'lumotlar Tahlili va Statistika*, 7(1), 34-48.
11. Zokirov, I. (2018). *Mashina o'rganish va statistika: Yangi metodologiyalar va yondashuvlar*. *O'zbekiston Iqtisodiy Tahlil Jurnali*, 15(2), 89-103.
12. Anderson, C. (2012). *The rise of big data*. *Wired Magazine*, 20(4), 22-29.
13. Baesens, B. (2014). *Analytics in a big data world: The essential guide to data science and its applications*. Wiley.
14. Bell, S., & Vose, D. (2019). *Introduction to statistics and data analysis*. Elsevier.
15. Breiman, L. (2001). *Random forests*. *Machine Learning*, 45(1), 5-32.

