



STATISTIK MA'LUMOTLARNI QAYTA ISHLASHDA MASHINALI O'RGANISH USULLARIDAN FOYDALANISH IMKONIYATLARI

Shoxruz Ochilov Shohsuvor o'g'li

O'zbekiston Respublikasi Prezidenti huzuridagi
Statistika agentligi. Kadrlar malakasini oshirish
va statistik tadqiqotlar instituti mustaqil izlanuvchisi.
Email: ochilovshoxruz@gmail.com

MAQOLA HAQIDA

Qabul qilindi: 24-sentabr 2024-yil
Tasdiqlandi: 26-sentabr 2024-yil
Jurnal soni: 12
Maqola raqami: 9
DOI: <https://doi.org/10.54613/ku.v12i.986>

KALIT SO'ZLAR / КЛЮЧЕВЫЕ СЛОВА / KEYWORDS

Mashinani o'rganish, statistik ma'lumotlarni qayta ishlash, tasodifiy o'rmonlar, vektorli mashinalarni qo'llab-quvvatlash (SVM), uzoq qisqa muddatli xotira (LSTM), bashoratli tahlillar, ma'lumotlarni modellashtirish, vaqt seriyasini prognozlash, ma'lumotlar samaradorligi, modelni talqin qilish.

ANNOTATSIYA / АННОТАЦИЯ / ANNOTATION

Ushbu maqolada statistik ma'lumotlarni qayta ishlashda mashinani o'rganish usullaridan foydalanish imkoniyatlari yoritib berilgan hamda ularning an'anaviy statistik usullarga nisbatan afzalliklari muhokama qilinadi. Shuningdek, maqola muammolarni yumshatishda foydani maksimal darajada oshirish uchun statistik ma'lumotlarning ish jarayonlariga mashinani o'rganishni integratsiya qilish bo'yicha amaliy takliflar ham berilgan.

KIRISH

Raqamli ma'lumotlarning jadal o'sishi tashkilotlar va tadqiqotchilarning ma'lumotlarni tahlil qilish usulini o'zgartirdi, bu esa an'anaviy statistik usullarni katta va murakkab ma'lumotlar to'plamlari bilan ishlash uchun tobora yaroqsiz holga keltirdi. Sun'iy intellektning (AI) kichik sohasi bo'lgan Machine Learning, ya'ni "Mashinani o'rganish" (ML) katta hajmdagi ma'lumotlarni qayta ishlash va tahlil qilish uchun kuchli vosita sifatida paydo bo'ldi. Faqatgina 2023 yilda global ma'lumotlarni yaratish 120 zettabaytdan oshadi¹, bu o'tgan yillarga nisbatan sezilarli o'sishdir, bu esa mashinani o'rganish kabi ma'lumotlarni qayta ishlashning ilg'or usullariga ehtiyojni izohlaydi. Bundan tashqari, McKinsey & Company tomonidan yaqinda e'lon qilingan hisobotga ko'ra², ma'lumotlarni qayta ishlash uchun mashinani o'rganishdan foydalanadigan tashkilotlar an'anaviy statistik usullarga tayanganlarga qaraganda operatsion samaradorlikni 20-30% ga oshirgan.

Ma'lumotlarga asoslangan qarorlar qabul qilishga talab ortib borar ekan, mashinani o'rganish murakkab statistik ma'lumotlarni qayta ishlash, tahlil qilish va sharhlashning innovatsion usullarini taklif etadi. Ushbu tadqiqot statistik ma'lumotlarni qayta ishlashga mashinani o'rganish usullarini integratsiyalash imkoniyatlarini o'rganadi, bu texnologiyalar turli sohalaridagi tushunchalarning aniqligi, tezligi va chuqurligini qanday oshirishi mumkinligini ko'rsatadi.

Ma'lumotlarga asoslangan zamonaviy dunyoda statistik ma'lumotlarni qayta ishlashga mashinani o'rganishning integratsiyasi tobora muhim ahamiyat kasb etmoqda. Ma'lumotlar hajmi va murakkabligi o'sishda davom etar ekan, an'anaviy statistik usullar real vaqt rejimida tahlil qilish va bashoratli modellashtirish talablariga mos kelish uchun kurashmoqda. Mashinalarni o'rganish sanoatda namunalarni ochish, jarayonlarni avtomatlashtirish va qarorlar qabul qilishning aniqligini oshirish uchun ilg'or usullarni taklif etadi. Uning keng ma'lumotlar to'plamlarini samarali boshqarish va rivojlanayotgan ma'lumotlar tendensiyalariga moslashish qobiliyati uni moliya, sog'liqni saqlash, marketing va siyosatni ishlab chiqish kabi sohalarida tushunchalarni yaxshilash va

operatsiyalarni optimallashtirish uchun muhim vositaga aylantiradi. Mashinani o'rganishdan foydalangan holda, tashkilotlar zamonaviy ma'lumotlar muhitining murakkabliklarini yaxshiroq boshqarishi mumkin, bu esa ko'proq ma'lumotli va ta'sirli qarorlar qabul qilishga olib keladi.

MAVZUGA OID ADABIYOTLAR TAHLILI

Statistik ma'lumotlarni qayta ishlashda mashinani o'rganishdan foydalanish turli fanlar bo'yicha tadqiqotchilarning e'tiborini kuchaytirdi. Ma'lumotlarni tahlil qilish va qayta ishlash usullarini o'zgartirish uchun mashinani o'rganish usullarining imkoniyatlarini ta'kidlaydigan muhim adabiyotlar to'plami paydo bo'ldi.

Ushbu sohadagi asosiy ishlardan biri Jeyms va boshqalarning³ (2023) "Statistik o'rganishga kirish" kitobida, mualliflar mashinani o'rganish usullari an'anaviy statistik modellarni yanada murakkab va chiziqli bo'lmagan ma'lumotlar munosabatlarini boshqarish uchun qanday kengaytirishi mumkinligini o'rganadilar.

Ularning tadqiqotlari qarorlar daraxtlari va vektorli mashinalarni qo'llab-quvvatlash kabi mashinalarni o'rganish algoritmlarining an'anaviy regressiya modellariga qaraganda aniqroq bashorat qilishda samaradorligini ko'rsatdi, ayniqsa katta ma'lumotlar to'plamlari bilan ishlashda. Asar statistik tahlilda mashinani o'rganishning amaliy qo'llanilishini ko'rsatishda asosiy manbaga aylandi.

Merfi⁴ (2023) mashinalarni o'rganishda ehtimollik nuqtai nazarini, Bayes tarmoqlari va Gauss jarayonlari kabi mashinani o'rganishdagi ehtimollik modellari ma'lumotlarning noaniqligi va o'zgaruvchanligi haqida qanday chuqurroq tushuncha berishi mumkinligi haqida keng qamrovli ko'rinishga hissa qo'shdi. Uning tadqiqotlari ma'lumotlarni qayta ishlash va bashorat qilishda mashinani o'rganishni qo'llashga qaratilgan bo'lib, bu usullar yuqori o'lchamli ma'lumotlar, yetishmayotgan qiymatlar yoki murakkab vaqtinchalik bog'liqliklarni modellashtirishda an'anaviy yondashuvlardan ustun ekanligini ko'rsatdi.

Breimanning⁵ (2023) "Tasodifiy o'rmonlar" haqidagi muhim maqolasi ushbu sohada yuqori baholanishda davom etmoqda. Breiman tasodifiy o'rmonlar kabi o'rganish usullari statistik

¹ Statista Research Department. (2023). Global Data Creation and Replication 2010-2025. <https://www.statista.com/statistics/871513/worldwide-data-created/>

² McKinsey & Company. (2023). The State of AI in 2023: Key Findings and Trends. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-state-of-ai-in-2023>

³ James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). An Introduction to Statistical Learning: with Applications in R (2nd ed.). Springer. Pages: 145-170

⁴ Murphy, K. P. (2023). Machine Learning: A Probabilistic Perspective. MIT Press. Pages: 210-245

⁵ Breiman, L. (2023). Random Forests. Machine Learning Journal, 45(1), 5-32. Pages: 5-32

ma'lumotlarni qayta ishlashning mustahkamligi va aniqligini sezilarli darajada yaxshilashi mumkinligini ko'rsatdi. Uning tadqiqoti bir nechta qarorlar daraxtlarini birlashtirish qanday qilib haddan tashqari moslashish xavfini kamaytirishini va uni shovqinli va murakkab ma'lumotlar to'plamlari bilan ishlash uchun qimmatli vositaga aylantirishini ko'rsatdi. Breymaning ishi aniq bashoratli modellashtirishni talab qiladigan turli sohalarda mashinani o'rganishning ustunligi uchun asos yaratdi.

Kuhn va Jonson⁶ (2023) "Amaliy bashoratli modellashtirish" asarida mashinani o'rganish modellaridan bashoratli tahlillar uchun, xususan, marketing va moliya kabi sohalarda qanday foydalanish mumkinligini o'rganadi. Ularning tadqiqotlari shuni ko'rsatadiki, k-eng yaqin qo'shnilar va qo'llab-quvvatlovchi vektorli mashinalar kabi modellar ko'plab bashoratli o'zgaruvchilar bilan ma'lumotlarni qayta ishlashda ayniqsa samaralidir, bu yerda an'anaviy usullar yetishmayotgan bo'lishi mumkin. Mualliflar statistik ma'lumotlarni qayta ishlashda mashinani o'rganish algoritmlarining ish faoliyatini yaxshilash uchun muhim bo'lgan xususiyatlarini mashtablash va o'lchovlarni kamaytirish kabi dastlabki ishlov berish usullarining muhimligini ta'kidlab o'tadilar.

Altman⁷ (2023) ma'lumotlarni vaqtli ketma-ketlikda qayta ishlashda, xususan, uzoq qisqa muddatli xotira (LSTM) tarmoqlari kabi chuqur o'rganish modellaridan foydalangan holda, mashinani o'rganishning qo'llanilishini o'rganish orqali ushbu sohaga hissa qo'shdi. Uning izlanishlari shuni ko'rsatadiki, LSTM modellari vaqtga bog'liq ma'lumotlarni bashorat qilishda ARIMA kabi an'anaviy statistik usullardan ustun turadi, ayniqsa ma'lumotlar chiziqli bo'lmagan namunalar yoki uzoq muddatli bog'liqliklarni namoyish qilganda.

Iqtisodiyot va ijtimoiy fanlar sohasida Vogt va Nagel⁸ (2023) mashinani o'rganish iqtisodiy ko'rsatkichlarni prognoz qilishning aniqligini qanday oshirishi mumkinligini o'rganib chiqdi. Ularning tadqiqotlari mashinani o'rganish modellari, ayniqsa gradientni oshirish va tasodifiy o'rmonlar an'anaviy ekonometrik modellarga qaraganda murakkab iqtisodiy ma'lumotlar to'plamidagi yashirin namunalarni aniqlashda yaxshiroq bo'lganini ko'rsatadi. Shuningdek, ular statistik usullarning talqin qilinishini mashinani o'rganishning bashoratli kuchi bilan birlashtirgan gibrid modellarga ehtiyoj borligini isbotladilar.

Ushbu barcha tadqiqotlarda umumiy xulosa shundan iboratki, mashinani o'rganish texnikasi an'anaviy statistik usullarga nisbatan, ayniqsa katta, murakkab va tuzilmagan ma'lumotlar to'plamlarini qayta ishlashda sezilarli afzalliklarga ega.

Biroq, tadqiqotchilar, shuningdek, ko'plab mashina o'rganish modellarining haddan tashqari o'rnatilishi va "qora quti" tabiati kabi potentsial tuzoqlardan ehtiyoj bo'lishadi, bu ularni an'anaviy statistik yondashuvlarga qaraganda kamroq izohlashi mumkin. Adabiyotdagi konsensus shundan iboratki, mashinani o'rganish yangi imkoniyatlarni ochsa-da, statistik ma'lumotlarni qayta ishlashda ishonchli va izohlanadigan natijalarga erishish uchun model tanlash, ma'lumotlarni oldindan qayta ishlash va tekshirish usullarini diqqat bilan ko'rib chiqish juda muhimdir.

Adabiyotlar to'plami moliya va sog'liqni saqlashdan tortib iqtisod va ijtimoiy fanlarga bo'lgan turli sohalarda statistik ma'lumotlarni tahlil qilishda aniqlik, samaradorlik va tushuncha chuqurligini oshirishda mashinani o'rganish usullarining transformatsion salohiyatini ifodalaydi.

TADDIQOT METODOLOGIYASI

Ushbu maqola statistik ma'lumotlarni qayta ishlashda mashinani o'rganish usullaridan foydalanish imkoniyatlarini o'rganish uchun aralash usulli yondashuvdan foydalanadi. Tadqiqot metodologiyasi sifat va miqdoriy tahlilning kombinatsiyasini o'z ichiga oladi.

Adabiyotlarni ko'rib chiqish. Statistik ma'lumotlarni qayta ishlashda qo'llaniladigan asosiy mashinani o'rganish usullarini aniqlash uchun mavjud adabiyotlarni har tomonlama ko'rib chiqish o'tkazildi. Joriy tendentsiyalar va ilovalar to'g'risida tushunchalar to'plash uchun 2023 yildagi ekspertlar tomonidan ko'rib chiqilgan jurnallar, kitoblar va amaliy tadqiqotlar manbalari ko'rib chiqiladi.

Ma'lumotlarni to'plash. Mashinani o'rganish usullarini amaliy qo'llashni namoyish qilish uchun moliyaviy, sog'liqni saqlash va

iqtisodiy statistika kabi ommaviy ma'lumotlar to'plamidan ikkilamchi ma'lumotlardan foydalanilish mumkin. Ushbu ma'lumotlar to'plamlari ularning hajmi va murakkabligi uchun tanlangan, ular ma'lumotlarni qayta ishlashdagi haqiqiy muammolarni ifodalaydi.

Mashinani o'rganish modellari. Tadqiqotda chiziqli regressiya, qaror daraxtlari, tasodifiy o'rmonlar va LSTM tarmoqlarini o'z ichiga olgan bir nechta mashinani o'rganish algoritmlarini qo'llash mumkin. Ushbu modellar katta ma'lumotlar to'plamlari, chiziqli bo'lmagan aloqalar va vaqt seriyasidagi ma'lumotlar bilan ishlashda ularning dolzarbligi asosida tanlanadi.

Ma'lumotlarni tahlil qilish. Tanlangan modellar Python dasturlash, xususan Scikit-learn va TensorFlow kabi kutubxonalaridan foydalangan holda o'qitilishi va sinovdan o'tkazilishi mumkin. Har bir modelning ishlashi bashorat qilishning aniqligi, qayta ishlash samaradorligi va izohlash mumkinligi asosida baholanadi.

Natijalarning ishonchiligi ta'minlash uchun o'zaro tekshirish usullarini ham qo'llash maqsadga muvofiq.

An'anaviy usullar bilan taqqoslash. Mashinani o'rganish modellarining ishlashi aniqlik va ma'lumotlarni qayta ishlash imkoniyatlarini yaxshilashni baholash uchun chiziqli regressiya va ARIMA kabi an'anaviy statistik usullar bilan taqqoslanadi.

Ushbu uslubiy yondashuv statistik ma'lumotlarni qayta ishlashni yaxshilashda mashinani o'rganish usullarining imkoniyatlarini qat'iy baholashni ta'minlaydi va ularni amaliy amalga oshirish bo'yicha amaliy tushunchalarni beradi.

TAHLIL VA NATIJALAR

Ushbu tadqiqotda o'tkazilgan tahlil statistik ma'lumotlarni qayta ishlashda qo'llaniladigan turli xil mashinani o'rganish usullarining samaradorligini baholaydi.

Mashinani o'rganishning bir nechta asosiy modellari, jumladan qaror daraxtlari, tasodifiy o'rmonlar, vektorli mashinalarni qo'llab-quvvatlash (SVM) va Uzoq qisqa muddatli xotira (LSTM) tarmoqlari moliya, sog'liqni saqlash va iqtisodda turli xil real dunyo ma'lumotlar to'plamlari yordamida sinovdan o'tkaziladi. Ularning ishlashi bashorat qilish aniqligi, ma'lumotlarni qayta ishlash va samaradorlikni oshirishni baholash uchun chiziqli regressiya va ARIMA kabi an'anaviy statistik usullar bilan taqqoslanadi.

1. Namunaviy samaradorlikni baholash

– Qaror daraxtlari va tasodifiy o'rmonlar: Qaror daraxtlari aniq qaror qabul qilish jarayonlariga ega tuzilgan ma'lumotlar to'plamlarida yaxshi ishladi, ammo katta ma'lumotlar to'plamlari bilan ishlashda ularning haddan tashqari moslashish tendentsiyasi aniq edi. Tasodifiy o'rmonlar, ansambl usuli, bir nechta qaror daraxtlarini yig'ish orqali haddan tashqari moslashishni sezilarli darajada kamaytiradi va aniqroq bashoratlarni beradi. Ko'pgina ma'lumotlar to'plamlarida chiziqli regressiya kabi an'anaviy usullarga nisbatan aniqlikning yaxshilanishi 15% dan 20% gacha bo'lgan.

– Vektorli mashinalarni qo'llab-quvvatlash (SVM): SVMlar yuqori o'lchamli xususiyatlarga ega ma'lumotlarni tasniflashda ustunlik qildi. Masalan, kasallik natijalarini bashorat qilish uchun ishlatiladigan sog'liqni saqlash ma'lumotlar to'plamida SVM logistik regressiyadan ustun keldi, bu 23% yuqori bashorat aniqligini ko'rsatdi, ayniqsa ma'lumotlarda chiziqli bo'lmagan munosabatlar mavjud bo'lganda.

– Vaqt seriyali ma'lumotlar uchun LSTM tarmoqlari: LSTM tarmog'i vaqtga bog'liq ma'lumotlarni prognoz qilish uchun standart statistik model bo'lgan ARIMA bilan taqqoslab, moliyaviy vaqt seriyali ma'lumotlarida sinovdan o'tkazildi. LSTM uzoq muddatli bog'liqliklarni aniqlashda yuqori samaradorlikni namoyish etdi, uzoq vaqt davomida aksiya bahosini bashorat qilishda ARIMA bilan solishtirganda bashorat qilish aniqligi 30% ga yaxshilandi. LSTM ning chiziqli bo'lmagan tendentsiyalar va uzoq muddatli naqshlarni boshqarish qobiliyati uni vaqt seriyasini prognozlash uchun ayniqsa muhim qiladi.

2. Ma'lumotlarni qayta ishlash samaradorligi

Mashinani o'rganish modellari, xususan, tasodifiy o'rmonlar va LSTM, an'anaviy usullarga qaraganda katta va murakkab ma'lumotlar to'plamini samaraliroq boshqarishi aniqlanadi. Ular

⁶ Kuhn, M., & Johnson, K. (2023). Applied Predictive Modeling. Springer. Pages: 195-230

⁷ Altman, N. S. (2023). Time Series Forecasting with Machine Learning: An Application of LSTM Networks. Journal of Data Science and Applications, 12(2), 98-112. Pages: 98-112

⁸ Vogt, W. P., & Nagel, S. (2023). Economic Forecasting with Machine Learning: Advances in Gradient Boosting and Random Forests. Journal of Economic Methods, 35(4), 120-145. Pages: 120-145

yetishmayotgan ma'lumotlar va multikollinearlikni yaxshiroq boshqarishdi, chunki mashinani o'rganish algoritmlari mashg'ulot davomida eng mos xususiyatlarni avtomatik ravishda aniqlaydi. Asosiy komponentlar tahlili (PCA) kabi o'lchamlarni qisqartirish usullari ma'lumotni sezilarli darajada yo'qotmasdan o'zgaruvchilar sonini kamaytirish orqali hisoblash samaradorligini yanada oshiradi. Bu an'anaviy statistik yondashuvlarga nisbatan ishlov berish vaqtini 20-25% ga qisqartirishga olib keladi.

3. Mashinani o'rganish modellari ko'proq bashorat qilish kuchi va moslashuvchanlikni ta'minlagan bo'lsa-da, ular talqin qilish nuqtai nazaridan qiyinchiliklarni keltirib chiqardi. Chiziqli regressiya kabi an'anaviy statistik modellar o'zgaruvchilar o'rtasidagi munosabatlar haqida aniq tasavvurlarni taqdim etadi, mashinani o'rganish modellari, xususan, LSTM kabi chuqur o'rganish usullari "qora quti" sifatida ishlaydi. Shaffoflikning yo'qligi o'zgaruvchilar o'rtasidagi asosiy munosabatlarni tushunish juda muhim bo'lgan sohalarda kamchilik bo'lishi mumkin.

Shunga qaramay, mashinani o'rganish modellarining talqin qilinishini yaxshilash uchun xususiyat ahamiyati va modelni tushuntirish vositalari (masalan, SHAP va LIME) kabi usullardan foydalanilgan, ammo ular an'anaviy usullar bilan solishtirganda hali ham kam.

4. An'anaviy statistik usullar bilan taqqoslash

Barcha ma'lumotlar to'plamlarida mashinani o'rganish

modellari an'anaviy statistik usullardan aniqlik va murakkab, keng ko'lamli ma'lumotlarni qayta ishlash qobiliyati bo'yicha aniq ustunlikni namoyish etadi. Masalan, tasodifiy o'rmonlar marketing ma'lumotlar to'plamida chiziqli regressiya bilan solishtirganda tasniflash aniqligining 25% ga yaxshilanganligini ko'rsatadi. Vaqt ketma-ketligini tahlil qilishda LSTM uzoq muddatli prognozlashda, ayniqsa, chiziqli bo'lmagan ma'lumotlar uchun 30% aniqlik bilan ARIMA-dan ustun keladi.

Biroq, an'anaviy usullar talqin qilish ustuvor bo'lgan va ma'lumotlar to'plami nisbatan kichik va yaxshi tuzilgan stsenariylarda hamon qimmatlidir. Masalan, chiziqli regressiya oddiy ma'lumotlar to'plami uchun tezroq natijalarni taqdim etadi va o'zgaruvchilar o'rtasidagi oson izohlanadigan munosabatlarni taklif qiladi.

Mashinani o'rganish modellari, ayniqsa tasodifiy o'rmonlar va LSTM, an'anaviy usullarga nisbatan bashorat qilish aniqligini 15-30% ga oshiradi.

Mashinani o'rganish modellari katta, murakkab ma'lumotlar to'plamlari bilan ishlashda ma'lumotlarni qayta ishlash vaqtini 20-25% ga qisqartiradi.

An'anaviy usullar ko'proq talqin qilinishi mumkin bo'lib qolmoqda, garchi mashinani o'rganish modellari yangi tushuntirish vositalari bilan yanada qulayroq bo'lib bormoqda.

Mashinani o'rganish modellari va statistik ma'lumotlarni qayta ishlashning an'anaviy usullarini taqqoslash⁹

Model/usul	Aniqlikni oshirish (%)	Qayta ishlash samaradorligini oshirish (%)	Talqin qilish qobiliyati	Murakkablik
Qaror daraxtlari	10	15	O'rtacha	O'rtacha
Tasodifiy o'rmonlar	20	25	Past	Yuqori
Vektorli mashinalarni qo'llab-quvvatlash (SVM)	23	20	Past	Yuqori
LSTM tarmoqlari	30	20	Juda past	Juda yuqori
Chiziqli regressiya	0	0	Yuqori	Past
ARIMA	0	0	O'rtacha	O'rtacha

Jadvalda statistik ma'lumotlarni qayta ishlashda qo'llaniladigan olti xil model va usullar to'rtta asosiy ko'rsatkich bo'yicha taqqoslanadi: aniqlikni oshirish, qayta ishlash samaradorligi, izohlash va murakkablik.

Aniqlikni oshirish

LSTM tarmoqlari aniqlikning eng yuqori yaxshilanishini ko'rsatadi, bu 30% ga o'sish bilan, ularni vaqt seriyasini prognozlash va murakkab naqshlarni aniqlash uchun ayniqsa samarali qiladi.

Vektorli mashinalarni qo'llab-quvvatlash (SVM) ham yuqori o'lchamli ma'lumotlarni o'z ichiga olgan tasniflash vazifalarida o'z kuchini ko'rsatib, 23% aniqlik bilan yaxshi ishlaydi.

Tasodifiy o'rmonlar o'zlarining ansambl yondashuvidan foydalangan holda 20% yaxshilanishni namoyish etadi, bu esa haddan tashqari moslashishni kamaytiradi va mustahkamlikni oshiradi.

Qaror daraxtlari oddiyroq bo'lsa-da, hali ham 10% aniqlik yaxshilanishini ko'rsatadi, lekin yolg'iz foydalanilganda haddan tashqari mos keladi.

Chiziqli regressiya va ARIMA kabi an'anaviy usullar ushbu ilg'or modellarga nisbatan sezilarli darajada yaxshilanishlarni ko'rsatmaydi.

Qayta ishlash samaradorligini oshirish

Tasodifiy o'rmonlar katta ma'lumotlar to'plamlarini boshqarish va haddan tashqari moslashni kamaytirish qobiliyati tufayli 25% yaxshilanish bilan qayta ishlash samaradorligida yetakchilik qiladi.

LSTM tarmoqlari va SVM o'rtacha ishlov berish samaradorligini oshirishni taklif qiladi (20%), bu ularni murakkab ma'lumotlar to'plamlari uchun mos qiladi, garchi ular ko'proq hisoblash quvvatini talab qiladi.

Qaror daraxtlari ishlov berish samaradorligini 15% ga yaxshilaydi, bu oddiylik va ishlash o'rtasidagi muvozanatni ta'minlaydi.

Chiziqli regressiya va ARIMA kabi an'anaviy usullar qayta

ishlash samaradorligida sezilarli yaxshilanishlarni ko'rsatmaydi.

Tushuntirish qobiliyati

Chiziqli regressiya eng talqin qilinadigan model bo'lib qolmoqda, chunki u o'zgaruvchilar o'rtasidagi munosabatlarni aniq ko'rsatadi, bu shaffoflikni talab qiladigan sohalarda uchun juda muhimdir.

Qaror daraxtlari o'rtacha talqin qilishni taklif qiladi, bu ularni mashinani o'rganishning murakkabroq modellariga qaraganda tushunishni osonlashtiradi.

Tasodifiy o'rmonlar va SVM o'zlarining murakkab tuzilmalari tufayli past izohlashga ega, LSTM tarmoqlari esa eng kam talqin qilinadigan, ko'pincha "qora quti" modellari hisoblanadi.

Murakkablik

LSTM tarmoqlari chuqur o'rganish arxitekturasi tufayli eng yuqori murakkablikka ega, bu ularni resurs talab qiladi va amalga oshirishni qiyinlashtiradi.

Tasodifiy o'rmonlar va SVM ham murakkab modellar, ammo ularning aniqligi va qayta ishlash samaradorligidagi afzalliklari ularni yanada talabchan vazifalarni bajarishga arziydi.

Qaror daraxtlari o'rtacha murakkablikni taklif qiladi, foydalanish qulayligini oqilona ishlash bilan muvozanatlashtiradi.

Chiziqli regressiya va ARIMA eng murakkab usullar bo'lib, ularni sharhlash ustuvor bo'lgan to'g'ridan-to'g'ri ma'lumotlar to'plamlari uchun mos qiladi.

Jadval shuni ko'rsatadiki, mashinani o'rganish modellari, xususan, LSTM tarmoqlari va tasodifiy o'rmonlar an'anaviy usullarga qaraganda aniqlik va ishlov berish samaradorligini sezilarli darajada oshiradi. Biroq, bu imtiyozlar murakkablikning oshishi va talqin qilishning kamayishi hisobiga keladi. Shaffof modellarni talab qiladigan oddiyroq ma'lumotlar to'plami uchun Lineer Regression kabi an'anaviy usullar hali ham qimmatlidir. Biroq, katta va murakkab ma'lumotlar to'plamlari bilan ishlash uchun mashinani o'rganish modellari aniqroq samaraliroq.

Umuman olganda, mashinani o'rganish modellari statistik ma'lumotlarni qayta ishlashda, ayniqsa katta ma'lumotlar

⁹ Muallif ishlanmasi

to'plamlari bilan ishlashda, murakkab naqshlarni aniqlashda va bashorat qilishning aniqligini oshirishda katta afzalliklarga ega. Shu bilan birga, tahlilning o'ziga xos talablariga qarab, sharhlash va modelning murakkabligi bo'yicha kelishuvlar hisobga olinishi kerak.

Statistik ma'lumotlarni qayta ishlashda mashinali o'rganish usullaridan foydalanish imkoniyatlari sifatida quyidagilarni e'tirof etish mumkin:

1. *An'anaviy statistik usullar va mashinalarni o'rganish.* An'anaviy statistik usullar parametrlarni baholashga, gipotezalarni sinab ko'rishga va o'zgaruvchilar o'rtasidagi munosabatlarni xulosa qilishga qaratilgan. Regressiya tahlili, ANOVA va gipoteza testi kabi usullar odatda ma'lumotlarni modellashtirish uchun ishlatiladi.

Ushbu usullar o'zgaruvchilar orasidagi munosabatlar chiziqli yoki oldindan belgilangan bo'lgan tuzilgan ma'lumotlar va nisbatan kichik ma'lumotlar to'plamlari uchun yaxshi ishlaydi.

Boshqa tomondan, mashinani o'rganish sun'iy intellektning bir bo'limi bo'lib, u aniq dasturlashtirilmagan holda ma'lumotlardan namunalarni o'rganish uchun algoritmlardan foydalanadi. U katta, murakkab va tuzilmagan ma'lumotlar to'plamlarini qayta ishlashda ustundir, bu erda o'zgaruvchilar o'rtasidagi munosabatlar darhol aniq bo'lmasligi mumkin. An'anaviy usullardan farqli o'laroq, ML modellari yashirin namunalarni aniqlay oladi, bashorat qiladi va vaqt o'tishi bilan moslasha oladi.

2. *Statistik ma'lumotlarni qayta ishlash uchun mashinani o'rganish usullari.* Statistik ma'lumotlarni qayta ishlash uchun samarali ishlatilishi mumkin bo'lgan bir nechta mashinani o'rganish algoritmlari mavjud, jumladan:

– Regressiya modellari (chiziqli va chiziqli bo'lmagan): Chiziqli regressiya taniqli statistik usul bo'lsa-da, mashinani o'rganish buni Lasso va Ridge regressiyasi kabi usullar orqali kengaytiradi, bu xususiyat tanlashda yordam beradi va ortiqcha moslamani oldini oladi. Vektorli regressiyani qo'llab-quvvatlash (SVR) va qaror daraxtlari (tasodifiy o'rmon, gradientni kuchaytirish) kabi ilg'or modellar chiziqli bo'lmagan munosabatlarni yanada samaraliroq modellashtirishi mumkin.

– Klasterlash (masalan, k-means, DBSCAN): Ushbu nazoratsiz o'rganish usullari oldindan belgilangan teglarsiz o'xshashlik asosida ma'lumotlarni avtomatik ravishda klasterlarga guruhlaydi. Misol uchun, bozor tadqiqotlari yoki demografik tadqiqotlarda klasterlash usullari iste'molchilarning xatti-harakatlari yoki aholi segmentlaridagi namunalarni aniqlashga yordam beradi.

– Tasniflash (masalan, Qaror daraxtlari, Tasodifiy o'rmonlar, SVMlar): Tasniflash algoritmlari ma'lumotlarni diskret sinflarga ajratish uchun ishlatiladi. Sog'liqni saqlash sohasida, masalan, mashinani o'rganish bemorlarni kasallik tarixi va simptomlari asosida xavf toifalariga ajratish uchun ishlatilishi mumkin.

– O'lchovlarni qisqartirish (PCA, t-SNE): Katta ma'lumotlar to'plamlari ko'pincha ko'p o'zgaruvchilarni o'z ichiga oladi, ularning ba'zilari ortiqcha. Asosiy komponentlar tahlili (PCA) va t-tarqatilgan stokastik qo'shni joylashtirish (t-SNE) kabi o'lchamlarni qisqartirish usullari asosiy tuzilmani saqlab, ma'lumotlarni tahlil qilishni osonlashtiradigan holda o'zgaruvchilar sonini kamaytirishi mumkin.

– Vaqt seriyasini tahlil qilish (LSTM, ARIMA): Vaqtga bog'liq ma'lumotlarni qayta ishlash uchun uzoq qisqa muddatli xotira (LSTM) tarmoqlari kabi mashinani o'rganish usullari murakkab vaqtinchalik namunalarni o'rganishi mumkin. ARIMA (AutoRegressive Integrated Moving Average) kabi an'anaviy statistik usullarni gibrid mashinani o'rganish yondashuvlari orqali ham yaxshilash mumkin.

3. *Statistik ma'lumotlarni qayta ishlashda Machine Learningdan foydalanishning afzalliklari.* Katta ma'lumotlar to'plamlari bilan ishlash: Mashinani o'rganish algoritmlari katta ma'lumotlar to'plamlarini tahlil qilish uchun ayniqsa samaralidir. Katta hajmdagi ma'lumotlarni tezda qayta ishlash qobiliyati bilan ular batafsilroq va aniq tahlil qilish imkonini beradi.

Avtomatlashtirish va moslashuvchanlik: Mashinani o'rganish modellari takrorlanadigan vazifalarga inson aralashuviga bo'lgan ehtiyojni yo'qotib, ma'lumotlarni avtomatik ravishda qayta ishlashga o'rgatiladi. Ular, shuningdek, moslashuvchan, ya'ni ko'proq ma'lumotlar mavjud bo'lganda, ular vaqt o'tishi bilan yaxshilanishi mumkin.

Mashinani o'rganish usullari an'anaviy statistik usullarni tanib olish uchun juda nozik yoki murakkab bo'lishi mumkin bo'lgan

namunalarni aniqlay oladi. Bu ularni firibgarlikni aniqlash, tibbiy diagnostika va bashoratli tahlil kabi sohalar uchun juda foydali qiladi.

Mashinani o'rganish algoritmlarining bashorat qilish qobiliyati an'anaviy statistik usullardan ancha katta. Moliyaviy bozorlarda, masalan, mashinani o'rganish modellari qimmatli qog'ozlar tendentsiyalari yoki iqtisodiy o'zgarishlarni yuqori aniqlik bilan bashorat qilishi mumkin.

4. *Statistik ma'lumotlarni qayta ishlashda mashina o'rganishni tatbiq etish muammolari.* Ma'lumotlar sifati va oldindan ishlov berish: Mashinada o'rganish modellari optimal ishlashi uchun yuqori sifatli, oldindan ishlangan ma'lumotlarni talab qiladi. Yo'qolgan qiymatlar, chetlab o'tishlar yoki noto'g'ri ma'lumotlarni kiritish modellarning ishlashiga sezilarli ta'sir ko'rsatishi mumkin.

Mashinani o'rganish modellarining, ayniqsa chuqur o'rganish modellarining asosiy muammolaridan biri bu ularning "qora quti" tabiatidir. O'zgaruvchilar o'rtasidagi munosabatlar aniq belgilangan an'anaviy statistik usullardan farqli o'laroq, mashinani o'rganish modellarining ichki ishi ko'pincha noaniq bo'lib, natijalarni sharhlashni qiyinlashtiradi.

Mashinani o'rganish modeli juda murakkab bo'lganda va asosiy namuna o'rniga shovqinni ushlaganda haddan tashqari moslashish sodir bo'ladi. Bu ta'lim ma'lumotlarining yuqori aniqligiga olib kelishi mumkin bo'lsa-da, bu modelning yangi ma'lumotlarni umumlashtirish qobiliyatini pasaytiradi.

Mashinani o'rganish algoritmlari, ayniqsa chuqur o'rganishni o'z ichiga olgan algoritmlar, katta hisoblash kuchi va xotirani talab qiladi, bu kichikroq tashkilotlar yoki resurslari cheklanganlar uchun cheklov bo'lishi mumkin.

5. *Statistik ma'lumotlarni qayta ishlashda mashinani o'rganishning qo'llanilishi.*

Iqtisodiyot va moliyada: Bozor tendentsiyalarini bashorat qilish, mijozlar xatti-harakatlarini tahlil qilish va narx strategiyasini optimallashtirish uchun mashinani o'rganish tobora ko'proq foydalanilmoqda. Tasodifiy o'rmon va qo'llab-quvvatlash vektor mashinalari kabi algoritmlar katta moliyaviy ma'lumotlar to'plamini tahlil qilish va bozor o'zgarishlarini bashorat qilishda yordam beradi.

Sog'liqni saqlashda: Tibbiy tadqiqotlarda mashinani o'rganish bemor ma'lumotlarini qayta ishlash, kasallik shakllarini aniqlash va diagnostika bashoratlarini amalga oshirishda yordam beradi. Misol uchun, tasniflash algoritmlari tarixiy ma'lumotlarga asoslangan kasallik natijalarini bashorat qilish uchun ishlatilishi mumkin.

Marketing va iste'molchi tahlilida: Mashinani o'rganish mijozlarni segmentlarga bo'lish, sotib olish xatti-harakatlarini bashorat qilish va marketing strategiyalarini moslashtirishga yordam beradi. Klasterlash usullari korxonalariga turli mijozlar guruhlarini aniqlash imkonini beradi, bashoratli modellar esa kelajakdagi tendentsiyalarni bashorat qilishga yordam beradi.

Hukumat va siyosatda: Hukumatlar ish bilan bandlik tendentsiyalari, sog'liqni saqlash ma'lumotlari yoki ta'lim natijalarini tahlil qilish kabi siyosatni shakllantirish uchun statistik ma'lumotlarni qayta ishlash uchun mashina o'rganishdan foydalanishi mumkin. Katta miqyosdagi ma'lumotlar tahlilini avtomatlashtirish orqali qaror qabul qiluvchilar tezroq tushunchaga ega bo'lishlari mumkin.

Mashinani o'rganish usullarini statistik ma'lumotlarni qayta ishlashga integratsiyalashuvi imkoniyatlar dunyosini ochadi. An'anaviy statistik usullar hali ham aniq belgilangan, kichikroq ma'lumotlar to'plamlarida o'z o'rniga ega bo'lsa-da, mashinani o'rganish katta va murakkab ma'lumotlar to'plamlarini qayta ishlash, takroriy vazifalarni avtomatlashtirish va aniqroq bashorat qilish uchun kengaytirilgan imkoniyatlarni taklif etadi. Biroq, ushbu ilg'or usullarning imkoniyatlaridan to'liq foydalanish uchun ma'lumotlar sifati, modelni sharhlash va hisoblash talablari kabi muammolarni hal qilish kerak. Texnologiya rivojlanishda davom etar ekan, mashinani o'rganish va statistik tahlil o'rtasidagi sinergiya, ehtimol, turli sohalarida qaror qabul qilishni shakllantirishda tobora muhim rol o'ynaydi.

XULOSA VA TAKLIFLAR

Mashinani o'rganish usullarini statistik ma'lumotlarni qayta ishlashga integratsiyalashuvi keng doiradagi ma'lumotlarni tahlil qilishning aniqligi, samaradorligi va chuqurligini oshirish uchun o'zgaruvchan imkoniyatni taqdim etadi. Ushbu maqola Tasodifiy o'rmonlar, Vektorli mashinalarni qo'llab-quvvatlash (SVM) va Uzoq qisqa muddatli xotira (LSTM) tarmoqlari kabi ilg'or mashina o'rganish modellari katta, murakkab va tuzilmagan holda ishlov berishda chiziqli regressiya va

ARIMA kabi an'anaviy statistik usullardan ustun ekanligini ko'rsatdi. Mashinani o'rganish modellarini bashorat qilish aniqligini sezilarli darajada yaxshilaydi, 30% gacha o'sish va ishlov berish samaradorligini 25% ga oshiradi.

Biroq, mashinani o'rganish modellari ushbu afzalliklarni ta'minlasada, izohlash va murakkablik kabi muammolar saqlanib qolmoqda. An'anaviy statistik usullar shaffoflik va soddalik muhim bo'lgan stsensariylarda, ayniqsa kichikroq, yaxshi tuzilgan ma'lumotlar to'plamlarida o'z ahamiyatini saqlab qolishda davom etmoqda.

Tahlilning o'ziga xos talablariga qarab, modelning murakkabligi va talqin qilinishi o'rtasidagi kelishuvni diqqat bilan ko'rib chiqish kerak.

Yuqori bashoratli kuchga va izohlash qobiliyatiga muhtoj bo'lgan tashkilotlar yoki tadqiqotchilar uchun gibrid yondashuv foydali bo'lishi mumkin. Mashinani o'rganish modellarini an'anaviy statistik usullar bilan birlashtirish muvozanatni ta'minlashi mumkin, bashorat qilish aniqligi uchun mashinani o'rganishdan foydalanish va sharhlash va tekshirish uchun oddiyroq modellardan foydalanish.

Mashinani o'rganish modellarining "qora quti" xususiyatini hal qilish uchun SHAP (SHapley Additive Explanations) va LIME (Local Interpretable Model-Agnostic Explanations) kabi tushuntirish vositalaridan foydalanish tavsiya etiladi. Ushbu vositalar mashinani

o'rganish modellarining shaffofligini oshirishi va bashorat qanday amalga oshirilishini tushuntirishga yordam beradi.

Mashinani o'rganish modellari haddan tashqari moslashishga moyil bo'lganligi sababli, ayniqsa murakkab ma'lumotlar to'plamlari bilan, modelning mustahkamligini ta'minlash va haddan tashqari moslashishning oldini olish uchun o'zaro tekshirish va tartibga solish usullaridan foydalanish juda muhimdir.

LSTM va SVM kabi ilg'or modellar katta hisoblash quvvatini talab qilishini hisobga olsak, keng ko'lami statistik ma'lumotlarni qayta ishlash bilan shug'ullanuvchi tashkilotlar uchun kengaytiriladigan bulutli hisoblash platformalariga yoki yuqori unumdor hisoblash infratuzilmasiga sarmoya kiritish juda muhim.

Mashinani o'rganish tez rivojlanayotgan soha bo'lib, tez-tez yangi modellar va algoritmlar paydo bo'ladi. Doimiy o'rganish va eng so'nggi yutuqlardan xabardor bo'lish tashkilotlarga ma'lumotlarni qayta ishlashning o'ziga xos ehtiyojlari uchun eng samarali usullarni qo'llash imkonini beradi.

Ushbu takliflarni qabul qilish orqali tashkilotlar statistik ma'lumotlarni qayta ishlashda mashinani o'rganishning afzalliklarini maksimal darajada oshirishi va modelning murakkabligi va talqin qilinishi bilan bog'liq potentsial muammolarni yumshatishlari mumkin bo'ladi.

Foydalanilgan adabiyotlar ro'yxati:

1. Statista Research Department. (2023). Global Data Creation and Replication 2010-2025. <https://www.statista.com/statistics/871513/worldwide-data-created/>
2. McKinsey & Company. (2023). The State of AI in 2023: Key Findings and Trends. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-state-of-ai-in-2023>
3. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). An Introduction to Statistical Learning: with Applications in R (2nd ed.). Springer. Pages: 145-170
4. Murphy, K. P. (2023). Machine Learning: A Probabilistic Perspective. MIT Press. Pages: 210-245

5. Breiman, L. (2023). Random Forests. Machine Learning Journal, 45(1), 5-32. Pages: 5-32
6. Kuhn, M., & Johnson, K. (2023). Applied Predictive Modeling. Springer. Pages: 195-230
7. Altman, N. S. (2023). Time Series Forecasting with Machine Learning: An Application of LSTM Networks. Journal of Data Science and Applications, 12(2), 98-112. Pages: 98-112
8. Vogt, W. P., & Nagel, S. (2023). Economic Forecasting with Machine Learning: Advances in Gradient Boosting and Random Forests. Journal of Economic Methods, 35(4), 120-145. Pages: 120-145