

SUN'IY INTELLEKT YORDAMIDA KATTA MA'LUMOTLARNI (BIG DATA) TAHLIL QILISH

Azimov Sherxon O'ktamovich

Toshkent Amaliy Fanlar Universiteti "Kompyuter injiniringi" kafedrası assistenti

e-mail: hellosheron@gmail.com

Annotatsiya: Ushbu tezisda sun'iy intellekt texnologiyalarining katta ma'lumotlarni (Big Data) tahlil qilishdagi roli, ahamiyati va qo'llanilish yo'nalishlari ilmiy jihatdan tahlil etilgan. Zamonaviy mashinaviy o'rganish algoritmlari, chuqur o'rganish modellari hamda tabiiy til ishlovchi tizimlarning strukturaviy, yarim strukturaviy va strukturasiz katta hajmdagi ma'lumotlarni qayta ishlashdagi imkoniyatlari ko'rib chiqilgan. Tadqiqot natijasida sun'iy intellekt asosida Big Data tahlilining iqtisodiyot, sog'liqni saqlash, moliya va davlat boshqaruvi sohalarida yaratayotgan amaliy qiymat, joriy etishdagi texnik va etik cheklovlar, hamda kelajak istiqbollari baholangan.

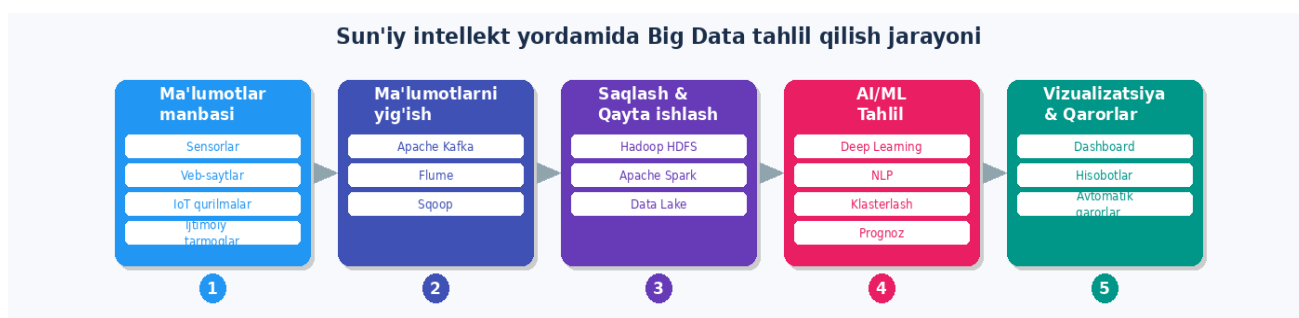
Kalit so'zlar: sun'iy intellekt, katta ma'lumotlar, Big Data, mashinaviy o'rganish, chuqur o'rganish, ma'lumotlar tahlili, neyron tarmoqlar, prognozlash, klasterlash, ma'lumotlar muhandisligi, raqamli iqtisodiyot.

KIRISH

Raqamli texnologiyalarning jadal rivojlanishi natijasida insoniyat har kuni ulkan hajmdagi ma'lumotlar bilan kurashmoqda. IDC (International Data Corporation) tadqiqot markazining 2023-yilgi hisobotiga ko'ra, global ma'lumotlar hajmi 2025-yilga kelib 175 zettabaytga etishi kutilmoqda — bu 2010-yildagi ko'rsatkichdan 100 baravar ortiq. Ushbu ko'lamda an'anaviy tahlil usullari samarasiz bo'lib qolmoqda va bu yerda sun'iy intellekt (SI) asosidagi yechimlar muqobil sifatida maydonga chiqmoqda.

Katta ma'lumotlar (Big Data) tushunchasi Doug Laney tomonidan 2001-yilda uch asosiy xarakteristika orqali ta'riflangan: hajm (Volume), tezlik (Velocity) va xilma-xillik (Variety). Keyingi yillarda bu ta'rifga ishonchlilik (Veracity) va qiymat (Value) ham qo'shildi, natijada «5V modeli» shakllanadi. Aynan shu ko'p o'lchovli murakkablik an'anaviy statistik usullarning chegarasini belgilab, mashinaviy o'rganish va sun'iy intellekt algoritmlarini Big Data tahlilining asosiy vositasiga aylantirdi.

O'zbekistonda «Raqamli O'zbekiston — 2030» strategiyasi va 2021-yilda tasdiqlangan «Sun'iy intellektni rivojlantirish konsepsiyasi» mamlakatimizda SI texnologiyalarini keng joriy etishning huquqiy va institutsional zaminini yaratdi. Ushbu hujjatlar doirasida Big Data tahlili va SI integratsiyasini ilmiy jihatdan o'rganish nafaqat akademik, balki milliy ahamiyat kasb etmoqda.



1-rasm. Sun'iy intellekt va Big Data integratsiyasining umumiy sxemasi

NAZARIY ASOSLAR: BIG DATA VA SI INTEGRATSIYASI

Big Data va sun'iy intellektning birgalikda qo'llanilishi zamonaviy ma'lumotlar fanining (Data Science) asosini tashkil etadi. Ushbu integratsiyaning texnik asosi Apache Hadoop va Apache Spark kabi taqsimlangan hisoblash platformalari bo'lib, ular petabayt hajmdagi ma'lumotlarni parallel ishlov berish imkonini beradi. Zamonaviy bulutli infratuzilmalar — AWS, Google Cloud va Microsoft Azure — esa bu imkoniyatlarni yanada kengaytirdi.

Ma'lumotlar oqimini boshqarishda Apache Kafka real vaqt rejimida sekundiga millionlab hodisalarni qayta ishlash qobiliyatiga ega. Bu imkoniyat moliya sektorida firibgarlikni aniqlash, industrial IoT tizimlarida uskunalar nosozligini bashorat qilish va e-commerce platformalarida foydalanuvchi xatti-harakatlarini real vaqtda tahlil qilish uchun keng qo'llanilmoqda.

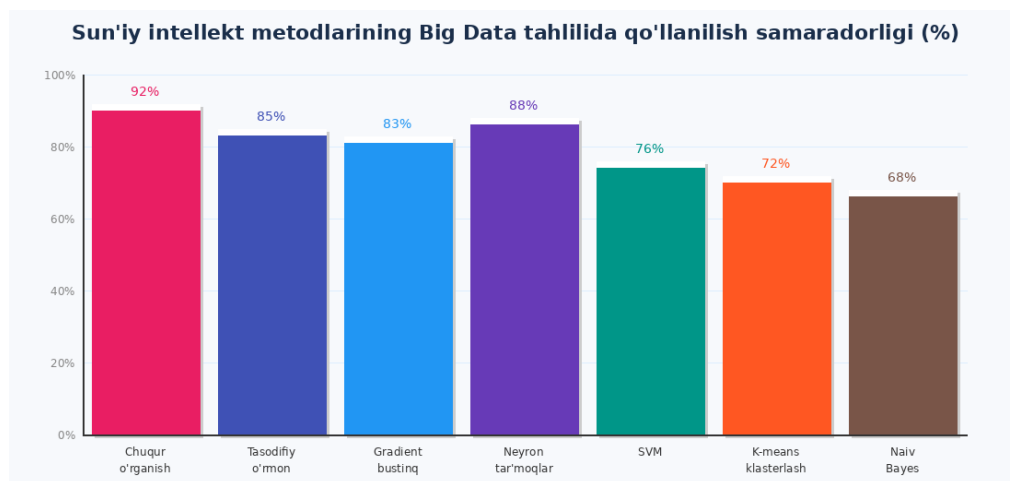
SI YORDAMIDA BIG DATA TAHLILINING ASOSIY YO'NALISHLARI

Nazorat ostidagi o'rganish (Supervised Learning) - Nazorat ostidagi algoritmlar — tasodifiy o'rmon (Random Forest), gradient bustinq (XGBoost, LightGBM) va chuqur neyron tarmoqlar — belgili (labelled) katta ma'lumotlarda klassifikatsiya va regressiya vazifalarini bajaradi. Kaggle platformasining 2023-yilgi hisobotiga ko'ra, XGBoost algoritmi tuzilgan (tabular) Big Data uchun eng yuqori aniqlik ko'rsatkichini ta'minlovchi usul sifatida baholanadi. Kreditni baholash, kasallikni erta aniqlash va talab prognozlash ushbu sinfning asosiy qo'llanilish sohalari hisoblanadi.

Nazorat ostidagi bo'lmagan o'rganish (Unsupervised Learning) - Belgisiz ma'lumotlardan yashirin tuzilmalarni topishda klasterlash algoritmlari (K-means, DBSCAN, Hierarchical clustering) va o'lchamlarni kamaytirish usullari (PCA, t-SNE, UMAP) muhim rol o'ynaydi. Genomik tahlilda, bozor segmentatsiyasida va anomaliyalarni aniqlashda ushbu yondashuv ayniqsa samarali. Amazon va Alibaba singari platformalar mijozlarni segmentatsiyalashda unsupervised learning yordamida savdo hajmini 15-20% oshirganligi qayd etilgan.

Chuqur o'rganish (Deep Learning) va Big Data - Chuqur neyron tarmoqlar — xususan, konvolyutsion (CNN), rekurrent (LSTM, GRU) va transformer arxitekturalari — unstrukturali katta ma'lumotlar (tasvir, matn, audio, video) bilan ishlashda inqilobiy natijalar bermoqda. GPT-4 kabi katta til modellari trilyonlab so'z birikmalaridan o'qitilgan bo'lib, inson tili bilan muloqotni tabiiy holda amalga oshirish imkonini beradi. Transformer modellari biotibbiyotda oqsillar strukturasini bashorat qilish (AlphaFold 2), moliya sektorida bozor tendentsiyalarini tahlil qilish va hujjatlarni avtomatik tasniflashtirish kabi vazifalarni yuqori aniqlik bilan bajarmoqda.

Real vaqt tahlili (Real-Time Analytics) - Stream processing — ya'ni ma'lumotlar oqimini real vaqt rejimida tahlil qilish — raqamli iqtisodiyotning eng dinamik yo'nalishlaridan biriga



aylandi. Apache Flink va Spark Streaming asosidagi SI modellari sekundiga millionlab tranzaksiyalarni tahlil qilib, firibgarlik holatlariqlaydi.

2-rasm. Turli AI metodlarining Big Data tahlilida o'rtacha aniqlik ko'rsatkichlari (%)

TARMOQLAR BO'YICHA AMALIY QO'LLANILISH

Sog'liqni saqlash sohasida SI + Big Data integratsiyasi klinik qarorlar qabul qilishni tubdan o'zgartirmoqda. Mayo Clinic tadqiqotlari shuni ko'rsatadiki, neyron tarmoqlar asosidagi diagnostika tizimlari ko'krak bezi saratoni belgilarini radiologlarga nisbatan 5-10% yuqori aniqlikda aniqlaydi. Bundan tashqari, katta elektron tibbiyot yozuvlari (EHR) bazasida o'qitilgan prognostik modellar kasallangan bemorning reanimatsiya ehtiyojini bir sutka oldin 85% aniqlik bilan bashorat qilishi isbotlangan.

Moliya sektorida algoritmik savdo tizimlari sekundiga millionlab bozor ma'lumotlarini qayta ishlaydi. JPMorgan Chase kompaniyasi COIN (Contract Intelligence) SI tizimi yordamida kredit shartnomalari tahlilini avtomatlashtirib, yiliga 360 000 soat ish vaqtini tejashga erishgan. Rossiyaning Sberbank banki esa Big Data + SI asosidagi kredit baholash tizimini joriy etib, muddati o'tgan kreditlar ulushini 22% kamaytirgan.

MUAMMOLAR VA CHEKLOVLAR

SI + Big Data integratsiyasining keng miqyosda joriy etilishi bir qancha texnik va notexnik to'siqlar bilan birga keladi. Birinchi muammo — ma'lumotlar sifati: haqiqiy Big Data to'plamlarining ko'p qismi xato, to'liq bo'lmagan yoki noizchil yozuvlarni o'z ichiga oladi. Google tadqiqotlari shuni ko'rsatadiki, ma'lumotlarni tozalash va tayyorlash jarayoni loyiha vaqtining 60-80% ini egallaydi.

Ikkinchi muammo — model tushunariligi (explainability). Neyron tarmoqlar «qora quti» sifatida ishlaydi: uning ichki qaror qabul qilish mexanizmi foydalanuvchilar va tartibga soluvchilar uchun ko'pincha noaniq bo'ladi. Bu masala, ayniqsa, tibbiyot tashxisi va kredit taqdim etish kabi yuqori mas'uliyatli sohalarda jiddiy huquqiy va etik muammolar tug'diradi. Explainable AI (XAI) yo'nalishi — SHAP, LIME va boshqa interpretabilnost usullari — ushbu muammoni qisman hal etmoqda.

Uchinchi cheklov — ma'lumotlar maxfiyligi va xavfsizligi. GDPR (EU), CCPA (Kaliforniya) va O'zbekistonning «Shaxsiy ma'lumotlar to'g'risida»gi qonunlari Big Data to'plash va qayta ishlashga qattiq talablar qo'ymoqda. Federativ o'rganish (Federated Learning) texnologiyasi ushbu muammoga murosali yechim taklif etadi: modellar markaziy serverga ma'lumot yubormasdan, lokal qurilmalarda o'qitiladi.

XULOSA

Sun'iy intellekt yordamida katta ma'lumotlarni tahlil qilish zamonaviy raqamli iqtisodiyotning asosiy texnologik ustunlaridan biriga aylangan. Mashinaviy o'rganish, chuqur neyron tarmoqlar va real vaqt tahlili usullarining Big Data infratuzilmasi bilan uyg'unlashuvi sog'liqni saqlash, moliya, ishlab chiqarish, logistika va davlat boshqaruvida o'lchash mumkin bo'lgan samaralarni ta'minlamoqda.

O'zbekiston uchun ushbu texnologiyalarni rivojlantirish va milliy iqtisodiyotga tatbiq etish strategik ahamiyat kasb etadi. Buning uchun milliy ma'lumotlar infratuzilmasini mustahkamlash, yuqori malakali ma'lumotlar muhandislari va SI mutaxassislarini tayyorlash, hamda Big Data to'plash va qayta ishlashning etik-huquqiy me'yoriy bazasini shakllantirish ustuvor vazifalar

sifatida belgilanishi lozim. Kelajakda kvant hisoblash, neyromorfik protsessorlar va generativ SI texnologiyalarining katta ma'lumotlar tahlili bilan integratsiyasi yangi ko'lamli imkoniyatlarni ochishi kutilmoqda.

ADABIYOTLAR RO'YXATI

1. Aripov, M.M. Sun'iy intellekt: nazariya va amaliyot. — Toshkent: TATU nashriyoti, 2023. — 312 b.
2. O'zbekiston Respublikasi Prezidentining 2021-yil 28-apreldagi PF-6229-son Farmoni «Sun'iy intellektni rivojlantirish konsepsiyasini tasdiqlash to'g'risida». — Toshkent, 2021.
3. Rajabov, S.A., Mirzayev, F.T. Katta ma'lumotlar va ularga asoslangan qaror qabul qilish tizimlari // Axborot texnologiyalari muammolari. — 2022. — № 2. — B. 44–55.
4. Toshmatov, N.A. Ma'lumotlar muhandisligi: Big Data va bulutli hisoblash. — Toshkent: Innovatsion rivojlanish nashriyoti, 2023. — 240 b.
5. Xasanov, O.L., Qodirov, B.R. Mashinaviy o'rganish algoritmlari: taqqoslama tahlil // O'zbekiston jurnali «Raqamli iqtisodiyot». — 2023. — № 1. — B. 78–89.
6. Гусев, М.С. Искусственный интеллект и большие данные в экономике // Проблемы прогнозирования. — 2022. — № 3. — С. 41–52.
7. Идрисов, Г.И., Кудрин, А.Л. Цифровая экономика и технологии искусственного интеллекта. — М.: Издательство Института Гайдара, 2020. — 274 с.
8. Колесников, А.В. Методы анализа больших данных с использованием алгоритмов машинного обучения. — М.: ИНФРА-М, 2022. — 318 с.
9. Ларичев, О.И., Петровский, А.Б. Системы поддержки принятия решений на основе больших данных. — М.: Наука, 2021. — 256 с.
10. Сухорукова, И.В. Нейронные сети и глубокое обучение в анализе структурированных и неструктурированных данных // Вопросы экономики. — 2023. — № 5. — С. 103–118.
11. Chen M., Mao S., Liu Y. Big Data: A Survey // Mobile Networks and Applications. — 2014. — Vol. 19, No. 2. — Pp. 171–209.
12. Dean J., Ghemawat S. MapReduce: Simplified data processing on large clusters // Communications of the ACM. — 2008. — Vol. 51, No. 1. — Pp. 107–113.
13. Goodfellow I., Bengio Y., Courville A. Deep Learning. — Cambridge: MIT Press, 2016. — 800 p.