

## 12-MAVZU: KATTA HAJMDAGI MA'LUMOTLAR BILAN ISHLASH

### Reja:

- 12.1. Katta hajmdagi ma'lumotlar (Big Data) qayta ishlash texnologiyalari
- 12.2. Big Data imkoniyatlaridan foydalanish
- 12.3. Korxona faoliyatida Big Data texnologiyalarini qo'llash.



Yangi texnologiyalar, qurilmalar va aloqa vositalarining tezkor rivojlanishi, ijtimoiy tarmoqlarning keng miqyosda yoyilishi ta'sirida insoniyat tarafidan ishlab chiqilayotgan ma'lumotlar miqdori keskin tarzda oshib bormoqda. Insoniyat yaralgan vaqtdan 2003-yilgacha to'plangan ma'lumotlar 5 milliard gigabayt miqdoriga teng deb hisoblanadi va bu ma'lumotlar diskka yozilgan holatda butun bir futbol maydonini egallashi mumkin bo'ladi.

Ushbu miqdordagi ma'lumotlar 2011-yilda har ikki kunda yaratilgan bo'lsa, 2013-yilga kelib, insoniyat tomonidan 5 milliard gigabayt ma'lumot har 10 daqiqada yaratilmoqda. Bu o'z navbatida, fanda yangi atama — katta hajmli ma'lumotlar (Big Data) tushunchasining yaralishiga sabab bo'ldi. Bu turdagi katta hajmli ma'lumotlar bilan ishlashda, hozirda yangidan-yangi algoritmlar va texnologiyalar ishlab chiqilmoqda va amaliyotda keng qo'llanilmoqda.

Atrofimizdagi milliardlab to'planib qolgan ma'lumotlarni yig'ish, qayta ishlash va ulardan samarali foydalanish orqali ma'lum holat yoki obyekt to'g'risida tarixiy, joriy va kelajak vaqt uchun muhim bo'lgan bilimlarga ega bo'lishimiz mumkin. Buning uchun dastlab, ushbu ma'lumotlarning manbasini aniqlash muhim amallardan biri hisoblanadi. «Katta hajmli ma'lumotlar» tushunchasi katta o'lchamdagi ishlarni bajara oladigan operatorlarga nisbatan

ishlatiladi. Bu asrlar davomida oʻrnatilgan tartibni oʻzgartirib yuboradi va bizning fundamental bilimlarimizni, qaror qabul qilishni shubha ostiga qoʻyadi. Haqiqiy evolyutsiya bu maʼlumotlarni chiqarib beradigan kompyuterlarda emas, balki, maʼlumotlarning oʻzlarida va biz ularni toʻgʻri qoʻllay olishimizdadir. Google bir kunda bir petabaytdan ortiq maʼlumotlarni qayta ishlaydi. Ushbu koʻrsatkich AQSH kongressi kutubxonasida nashr qilinadigan materiallardan 100 barobar koʻproq. Atiga 10 yil avval mavjud boʻlmagan kompaniya — Facebook, soatiga 10 milliondan ortiq yangi rasmlar joylanishi bilan faxrlanishi mumkin. 800 millionta Googlening YouTube xizmati foydalanuvchilari oyiga har soniyada bir soatdan koʻp videolar joylashtiradi.

Bizni oʻrab turgan maʼlumotlar haqiqiy hajmining jadalligini, oʻsishini hisoblashni koʻp odamlar baholashga harakat qilishgan. Har xil narsalarni hisoblashgani uchun, ular har xil muvaffaqiyatlarga erishishdi. Eng toʻliq tadqiqotni Annenberg nomidagi Janubiy Kaliforniya universiteti qoshidagi kommunikatsiyaga ixtisoslashgan maktabdan Martin Gilbert olib bordi. U hamma ishlab chiqilgan, saqlangan, yuborilgan narsalarni hisoblashga harakat qildi. Bu faqat kitoblar, rasmlar, elektron xatlar, raqamli musiqa va videolar emas, balki video oʻyinlar, qoʻngʻiroqlar va avtomobil navigatsion tizimlari hamdir. Undan tashqari u ommaviy axborot vositalari, televideniya va radioning auditoriya kattaliklarini ham hisobga olgan. Uning hisoblariga koʻra, 2007-yilda qariyb, 2,25 zettabayt maʼlumot yuborilgan va saqlangan. Bu koʻrsatkich 20 yil avvalgisidan (435 ekzabayt) 5 barobar koʻproq.



Agar ommaviy axborot vositalaridan tashqari saqlanayotgan ma'lumotlar ko'rib chiqilsa, qiziq tendensiyalar paydo bo'ladi. 2007-yilda 300 ekzabayt saqlangan ma'lumotlarning 7% analog formatda (qog'oz hujjatlar, kitoblar, fotosuratlar), qolganlari esa raqamli formatda bo'lgan. 2000-yilda bo'lsa raqamli formatda saqlanadigan ma'lumotlar soni dunyodagi saqlanadigan ma'lumotlarning atiga bir chorak qismini tashkil etar edi. Qolgan 3 chorak ma'lumotlar esa qog'oz hujjatlarda, plyonkalarda, plastinkalarda, magnit plastinkalarda va boshqa tashuvchilarda bo'lgan. 2013-yilga kelib, butun dunyoda saqlanayotgan ma'lumotlar 1.2 zettabaytni tashkil etdi va ulardan sonli bo'lmagan ma'lumotlar atiga 2 foizni tashkil etdi. Dunyoda katta ma'lumotlardan katta hajmdagi ma'lumotlarni tahlil qilish, ba'zi hollarda esa hammasini qayta ishlash mumkin. Hamma ma'lumotlardan foydalangan holda biz aniq natijaga ega bo'lishimiz mumkin va kichik hajmdagi ma'lumotlar chegaralangandagi kamchiliklarini ko'rishimiz mumkin. Tanlash yo'li bilan baholab bo'lmaydigan katta ma'lumotlar detallarning segmentlarini va toifalari haqida aniq tasavvurga ega bo'lish mumkin. Katta hajmdagi ma'lumotlarga etibor qilgan holda aniqlikka bo'lgan qarashlarimizni kamaytirishimiz mumkin. Yaqingacha, barcha sonli qurilmalar aniqlikka asoslangan edi. Bizda barcha elektron jadval ustunlari ma'lumotlar tizimi bazasi sonlari singari bizning talablarga mos ko'rinishda yozilgan degan tushuncha shakllangan edi. Bunday fikr usuli «kichik ma'lumotlar» ustida qo'l keladi. Iloji boricha, hamma yozilgan ma'lumotlarni aniqroq hisoblash uchun kam ko'rsatkichlar o'lchanar edi.

Aniqlik ma'lumotlarning qattiq tekshiruvini talab qiladi. U kichkina hajmdagi ma'lumotlar uchun mos keladi va ba'zi hollarda u juda kerak (masalan, chek yozish uchun bank hisobida yetarli miqdorda pul bor yoki yo'qligini tekshirish uchun). Lekin ma'lumotlar olamida 100% aniqlilik bo'lishi mumkin emas va kerak ham emas. Agar biz ma'lumotlarning ko'p qismi doim o'zgaradigan ma'lumotlar bilan ishlasak, aniqlilik ikkinchi o'ringa tushib qoladi. Katta hajmli ma'lumotlar — inson hayoti uchun dunyoni qurshab olish va hisoblashlardagi harakatlar borasidagi katta olg'a qadamdir. Katta hajmli ma'lumotlar haqida yangicha tushunchalarni shakllantish uchun 3 ta asosiy holatni sanab o'tsak bo'ladi, bu holatlar bir-biriga bog'langan va bir-birini to'ldirib turadi. 1-barcha ma'lumotlardan qismlab yoki statik tanlanmadan ularni analiz qilish. 2-aniqlikka putur yetkazgan holda betartiblik bilan ishlashga tayyor bo'lish. 3-fikrlashni o'zgartirish: qiyin sababiyatdan ko'ra korrelatsiyalarga ishonish.

Tanlanish tushunchasi minimum materialdan maksimum foydalisini tanlab olish, eng kam berilganlardan katta hajmli tahlil qilishni ko'zlashdir. Endi esa tanlanish eski tushuncha deb qabul qilamiz va katta berilgan ma'lumotlar bilan ishlaymiz. Betartiblik — katta ma'lumotlarning ichki tavsifi emas, balki ishimiz bog'langan obyektiv reallikdir. Material obyektlar singari, ma'lumotlar bahosining iste'mol qilinishi kichiklashmaydi. Ma'lumotlarni qayta va qayta ishlash mumkin. Ular, iqtisodchilar tili bilan aytganda «taqqoslanmaydigan» tovarlar qatoriga kiradi. Shuningdek, moddiy egaliklardan farqli o'laroq,

ma'lumot ishlatilishdan to'xtamaydi. Ma'lumotlarning asl bahosi — xuddi okeandagi aysberglardek. Bir qarashda faqat kichik qism ko'rinadi, ayni paytda qolgan qismi suv tagida yotadi, yani ma'lumotlar qimmatligini faqat yuzaki emas, balki har tomonlama qarash kerak. Katta hajmli ma'lumotlar atrofimiz haqida ancha aniqroq prognozlar bera olishni ta'minlaydi. Biz bunday prognozlarga tayyor bo'lmasligimiz mumkin. Chunki bizning dunyoqarashimiz va tizim muassasalarimiz bunday ko'p ma'lumotlarga emas, balki kam ma'lumotlarga moslashtirilgan.

Katta hajmli ma'lumotlar ularni tahlil qilgan, saqlagan va ularni ikkinchi bor ishlatganlarga yangiliklarni ochib beradi. Chunki saqlanishlar qiymati ham shu darajada tushib ketadi, tahlil vositalari esa, kuchliroq, o'lcham va masshtab yig'indisi kunma-kun, soatma-soat o'saveradi. Agar Internet xavfsizlik tizimini xavf ostida qoldirgan bo'lsa, katta ma'lumotlar uchun bu masalani yechish imkoni bormi? Bu ularning qorong'u tomoni emasmi? Yana shunday bir xavf borki, biz axborot boshqaruvchilar qurbonlari bo'lishga tavakkal qilayapmiz. Bunda biz ma'lumotlarga va ular to'g'risida chiqayotgan tahlillarga sig'inishni boshlaymiz va oxir-oqibatda, ularni haddan tashqari ko'p ishlatishni boshlaymiz. Agar ulardan oqilona foydalanilsa, katta ma'lumotlar ratsionanal masalalar yechimi uchun yaxshi vosita hisoblanadi. Agarda ulardan aql bilan foydalanilmasa, u insoniyatning retssesiyasiga hissa qo'shishi, ya'ni mijozlar va ishchilarga noqulayliklar yaratishi yoki odamlarga ziyon yetkazishi mumkin.

Dunyo katta ma'lumotlar asri tomon suzayotgani uchun, aholi o'sha «tektonik» silkinishga yaqin yurmoqda. Katta ma'lumotlar bizning fundamental tasavvurlarimizni, o'zgartirishga majbur qilmoqda. Ular bizning yashash tarzimizga kirib borayotgani sababli ularning qanday o'sishini va potensial ziyonini boshqarishni o'rganishimizga to'g'ri kelayapti. Vaholangki, bunga asrlar emas, bir necha yillar qolgan bo'lishi ham mumkin. Shaxsiy hayot va ma'lumotlar xavfsizligini ta'minlash barchamizning o'z zimmamizda. Bizga o'z majburiyatlarimiz va huquqlarimizni bilib yurish zimmamizga yuklanadi. Boshqalarning qiziqishlarini nazorat qilish, bulardan ziyon ko'rmaslik maqsadida qiyin algoritmlarni tuzish uchun yangi muassasalar va ekspertlar kerak bo'ladi. Gap o'zgarishlarga ko'nikishda emas, balki yangi o'zgartirishlar kiritilishidadir. Katta ma'lumotlar aqlbovar qilmaydigan masalalarni yecha oladigan texnologiya deb tushuniladi, lekin yangilari undan ham hayron qoldiradi. Katta ma'lumotlar hayotimizni, ishimizni, fikrlashimizni o'zgartirish imkoniga ega.