

**KATTA MA'LUMOTLARNI TAHLIL QILISH TEXNOLOGIYALARI
VA VOSITALARI**

Qo'chqorov Sherali Xasan o'g'li

Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti Jizzax filiali

4-boshqich talabasi

sheraliqochqorov157@gmail.com

Mamirov Javohir Shodiyor o'g'li

Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti Jizzax filiali

4-boshqich talabasi

javohirbek.mamirov08@gmail.com

Axmatov Jasur Forxod o'g'il

Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti Jizzax filiali

4-boshqich talabasi

jasuraxmatov681@gamil.com

Anotatsiya. Ush bu maqolada Katta ma'lumotlarni tahlil qilish texnologiyalari haqida so'z boradi. Bu maqola orqali katta ma'lumotlarni tahlil qilish texnologiyalari, katta ma'lumotlar tahlili qayerda ishlatilishi, tahlil qilishning afzalliklarini va tahlil qilish mobaynida duch keladigan qiyinchiliklarni ko'rib chiqamiz.

Kalit so'zlar. Tavsiflovchi tahlil, Bashoratli tahlil, Retseptiv tahlil, Tashxisli tahlil, Ma'lumotlarni tahlil qilish, Machine learning.

Kirish

Big data atamasi 2008-yilda dunyoga kelgan. Nature jurnali muharriri Klifford Linch dunyo ma'lumotlar hajmining juda tez sur'atda o'sishiga bag'ishlangan maxsus sonida big data atamasini qo'llagan. Biroq, katta ma'lumotlar avval ham bo'lgan. Mutaxassislarning fikricha, kuniga 100 gb dan ko'p ma'lumot tushadigan oqimlarga big data deb aytiladi. Katta ma'lumotlarni analiz qilish, inson his etish imkoniyatidan tashqarida bo'lgan qonuniylatlarni

aniqlashda yordam beradi. Bu esa kundalik hayotimizdagi barcha sohalar, hukumatni boshqarish, tibbiyot, telekommunikatsiya, moliya, transport, ishlab chiqarish va boshqa sohalarni yanada yaxshilash, ularning imkoniyatlarini oshirish, muommolarga muqobil yechimlar izlab topish imkonini yaratadi.

Tahlil qilishning to‘rtta asosiy usuli mavjud:

Tavsiflovchi tahlil (descriptive analytics) – eng keng tarqalgan usul bo‘lib, u «nima bo‘ldi?» degan savolga javob beradi hamda real vaqtda kelayotgan va tarixiy ma‘lumotlarni tahlil qiladi. Asosiy maqsad ma‘lum bir sohadagi muvaffaqiyat yoki muvaffaqiyatsizlik sabablari va uning muntazamligini aniqlash hamda ushbu ma‘lumotlardan samarali foydalanish.

Bashoratli tahlil (predictive analytics) – mavjud ma‘lumotlarga asoslanib, voqealarning eng ehtimoliy rivojlanishini bashorat qilishga yordam beradi. Buning uchun shunga o‘xshash xususiyatlar to‘plamiga ega bo‘lgan har qanday ob‘yekt yoki hodisalarga asoslangan tayyor shablonlardan foydalanadi.

Retseptiv tahlil (prescriptive analytics) – bashoratli tahlil bilan solishtirganda undan keyingi daraja. Big Data va zamonaviy texnologiyalar yordamida biznesdagi yoki boshqa faoliyatdagi muammoli nuqtalarni aniqlash va kelajakda qaysi yo‘l bilan ulardan qochish mumkinligini hisoblash mumkin.

Tashxisli tahlil (diagnostic analytics) – sodir bo‘lgan voqea sabablarini tahlil qilish uchun ma‘lumotlardan foydalanadi. Bu hodisalar va harakatlar o‘rtasidagi anomaliyalar va tasodifiy aloqalarni aniqlashga yordam beradi.

Big Data bilan ishlash usullari va vositalari Katta ma‘lumotlarni to‘plash va tahlil qilishning asosiy usullari quyidagilarni o‘z ichiga oladi:

- Data Mining - assotsiativ qoidalarni o‘qitish, tasniflash, klaster va regressiya tahlili;
- krod sourcing - bu inson yordamida ma‘lumotlarni toifalash va boyitish, ya‘ni, uchinchi shaxslarning ixtiyoriy yordami bilan;
- raqamli signalga ishlov berish va tabiiy tilda ishlov berish kabi ma‘lumotlarni aralashtirish va birlashtirish;

• sun'iy neyron tarmoqlari, tarmoqni tahlil qilish, optimallashtirish usullari va genetik algoritmlarni o'z ichiga olgan holda mashinani o'rganish;

- takrorlanishlarni aniqlash;
- bashoratli tahlil;
- simulyatsiya modellashtirish;
- mekansal va statistik tahlil;

• analitik ma'lumotlarni vizualizatsiya qilish - rasmlar, grafikalar, diagramma, jadvallar. Katta ma'lumotlar bilan ishlash uchun dasturiy va apparat vositalari kengaytirish, parallel hisoblash va tarqatishni ta'minlaydi, chunki doimiy o'sish katta ma'lumotlarning asosiy xususiyatlaridan biridir. Asosiy texnologiyalarga aloqador bo'lmagan ma'lumotlar bazasi (NoSQL), MapReduce ma'lumotlarini qayta ishlash modeli, Hadoop klasteri ekotizimining tarkibiy qismlari, R va Python dasturlash tillari, shuningdek Apache-ning ixtisoslashtirilgan mahsulotlari (Spark, AirFlow, Kafka, HBase va boshqalar) kiradi.

Katta ma'lumotlarning afzalliklari

Katta ma'lumotlar o'z vaqtida - Har bir ish kunining 60%, bilimlar ma'lumotlarini topish va boshqarish uchun sarflaydi.

Katta ma'lumotlarga kirish mumkin - Katta ma'murlarning yarmi to'g'ri ma'lumotlarga kirish qiyinligini ta'kidlamoqda.

Katta ma'lumotlar yaxlitdir - Ma'lumotlar hozirda tashkilot ichidagi siloslarda saqlanadi. Marketing ma'lumotlarini, masalan, veb-tahlil, mobil tahlil, ijtimoiy tahlil, CRMs, A/B test vositalari, elektron pochta marketing tizimlari va boshqalar... har biri o'z silosiga e'tibor qaratadi.

Katta ma'lumotlar ishonchli - Kompaniyalarning 29 foizi ma'lumotlar sifati past bo'lgan pul xarajatlarini o'lchaydilar. Mijozlar bilan aloqa ma'lumotlarini yangilash uchun bir nechta tizimlarni kuzatish kabi oddiy narsalar millionlab dollarlarni tejashga yordam beradi.

Katta ma'lumotlar dolzarbdir - Kompaniyalarning 43 foizi o'zlarining ahamiyatsiz ma'lumotlarni filtrlash vositalari qobiliyatidan norozi. Mijozlarni

veb-saytingizdan filtrlash kabi oddiy narsa Analytics sotib olish harakatlaringiz haqida bir tonna tushuncha berishi mumkin.

Katta ma'lumotlar xavfsizdir - Ma'lumotlar xavfsizligining o'rtacha buzilishi har bir mijoz uchun 214 dollarni tashkil qiladi. Katta ma'lumotli xosting va texnologik sheriklar tomonidan qurilgan xavfsiz infratuzilmalar o'rtacha kompaniyani yillik daromadining 1.6 foizini tejashga qodir.

Katta ma'lumotlar vakolatli - Tashkilotlarning 80% o'zlarining ma'lumot manbalariga qarab haqiqatning bir nechta versiyalari bilan kurashishadi. Ko'plab tekshirilgan manbalarni birlashtirib, ko'proq kompaniyalar yuqori aniqlikdagi razvedka manbalarini ishlab chiqarishlari mumkin.

Katta ma'lumotlar harakatga yaroqlidir - Eskirgan yoki yomon ma'lumotlar natijalariga ko'ra 46% kompaniyalar milliardlarga tushishi mumkin bo'lgan noto'g'ri qarorlar qabul qilishadi.

Katta ma'lumotlar tahlilining qiyinchiliklari

Katta ma'lumotlar tahlilining (Big Data) bir qancha qiyinchiliklari mavjud bo'lib, ular asosan quyidagi omillar bilan bog'liq:

Ma'lumot hajmi - Katta ma'lumotlar juda katta hajmdagi (terabaytlar yoki petabaytlar) ma'lumotlarni o'z ichiga oladi. Bunday hajmdagi ma'lumotlarni saqlash, qayta ishlash va tahlil qilish uchun juda katta hajmdagi resurslar, kuchli kompyuterlar va optimallashtirilgan algoritmlar talab qilinadi.

Ma'lumot xilma-xilligi - Ma'lumotlar tuzilmali (structured), tuzilmasiz (unstructured), va qisman tuzilmali (semi-structured) bo'lishi mumkin. Masalan, hujjatlar, videolar, sensorlar ma'lumotlari, loglar va ijtimoiy tarmoqlardagi postlar bir xil formatda bo'lmasligi mumkin. Bularni bir tizimda to'plash va tahlil qilish murakkab jarayon.

Aniqlik va sifat - Katta ma'lumotlar ko'pincha noto'g'ri, to'liq bo'lmagan yoki shovqinli bo'lishi mumkin. Ma'lumotlarning sifati past bo'lsa, noto'g'ri tahlillar va xulosalar chiqarilishi ehtimoli katta. Shuning uchun, to'g'ri ma'lumotlarni ajratib olish va noto'g'ri yoki ortiqcha ma'lumotlarni filtrlash katta ahamiyatga ega.

Ma'lumot xavfsizligi va maxfiyligi - Ko'p ma'lumotlarni yig'ish va saqlash ma'lumotlarning xavfsizligi va maxfiyligi masalasini keltirib chiqaradi. Ma'lumotlarga ruxsatsiz kirish yoki ularni buzish katta korporatsiyalar uchun xavf tug'diradi. Shu sababli, xavfsizlik choralari ko'rish katta e'tibor talab etadi.

Kadrlar yetishmovchiligi - Katta ma'lumotlar tahlili sohasida malakali mutaxassislarning yetishmasligi ham jiddiy qiyinchiliklardan biridir. Ma'lumotlar olimi (data scientist), ma'lumotlar muhandisi (data engineer) kabi kasblar uchun zarur bo'lgan bilimlar va ko'nikmalar keng qamrovli bo'lib, ushbu sohada yetuk kadrlarni topish qiyin bo'lishi mumkin.

Tahlil qilish texnikalari murakkabligi - Katta ma'lumotlarni tahlil qilish uchun mashinani o'qitish (machine learning), sun'iy intellekt (AI), va statistika kabi ilg'or usullar talab qilinadi. Ushbu texnologiyalarni samarali qo'llash uchun maxsus bilimlar va tajriba kerak bo'ladi.

Yuqoridagi qiyinchiliklar katta ma'lumotlarni samarali boshqarish va tahlil qilish jarayonida hal qilish zarur bo'lgan asosiy muammolarni tashkil qiladi.

Katta ma'lumotlar tahlilning Sohada qo'llanilishi

«Katta ma'lumotlar» tahlili kompaniyalarning asosini tashkil qilib, real vaqtda katta ma'lumotlar to'plamini boshqarish, qayta ishlash, optimallashtirishga yordam beradi va qaror qabul qilish qobiliyatini yaxshilaydi. Bundan tashqari, katta ma'lumotlar va biznes-tahlilning asosiy maqsadi kompaniyalarga o'z mijozlarini yaxshiroq tushunishga yordam berish va marketing kampaniyalarini yaxshilashdir.

Ba'zi tarmoqlar katta ma'lumotlardan foydalanishda katta yutuqlarga erishgan bo'lsa-da, boshqalari hali ham katta ma'lumotlarga yetarli e'tibor qaratishgani yo'q. 2025 yilga kelib bank sohasidagi katta ma'lumotlar tahlil bozori 62,10 mlrd dollargacha o'sishi mumkin. Bank sohasidagi katta ma'lumotlarning statistikasiga ko'ra, jahon bank sektori infratuzilmasi allaqachon katta ma'lumotlarni tahlil qilishni o'z ichiga oladi.

Foydalanilgan adabiyotlar:

1. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И., Методы и модели анализа данных: OLAP и Data Mining.СПб.: БХВ-Петербург, 2004. - 336 с.
2. Елманова Н., Федоров А. Введение в OLAP-технологии Microsoft. СПб.:БХВ-Петербург, 2014.-232 с.
3. Вячеслав Дюк., Дюк В.А., Самойленко А.П. Data Mining. Учебный курс СПб: Питер, 2001. -368 с.
4. Паклин Н.Б., Орешков В.И. Бизнес аналитика: от данных к знаниям. СПб Питер, 2012.-461 с.
5. Киселев М., Соломатин Е. Средства добычи знаний в бизнесе и финансах.Открытые системы. 1997. № 4. С. 41-44
6. John F. Elder IV & Dean W. Abbott.KDD-98: A Comparison of Leading Data Mining Tools. Fourth International Conference on Knowledge Discovery & Data Mining, August 28, 1998. New York
7. Damiaan Zwietering, Helena Gottschalk, Hosung Kim, Joerg Reinschmidt. Intelligent Miner for Data: Enhance Your Business Intelligence J. June 1999, International Technical Support Organization, SG 245422
8. Эделстейн Г. Интеллектуальные средства анализа, интерпретации и представления данных в информационных хранилищах ComputerWeekМосква. 1996. № 16. С. 32-33
9. А.Н. Горбань. Методы нейроинформатики. КГТУ, Красноярск, 1998. 205 с
10. А.Н. Горбань, А.Н. Кирдин и др, В.Л. Дунин-Барковский. Нейроинформатика. Новосибирск: Наука, 1998. 296с