

DATA MINING METODLARI VA BOSQICHLARI

Mirzakarimov Baxtiyor Abdusalomovich

“University of economics and pedagogy”NOTM,

“Iqtisodiyot” kafedrasi v/b professor +998905442766

baxtiyormirzakarimov19@gmail.com.

Annotatsiya: *Ushbu maqolada zamonaviy axborot texnologiyalari taraqqiyotining asosiy yo‘nalishlaridan biri bo‘lgan data mining, ya’ni ma’lumotlar konchilik texnologiyalarining nazariy asoslari, metodlari va amaliy qo‘llanilishi yoritiladi. Katta hajmdagi ma’lumotlar ichidan foydali bilimlarni ajratib olish jarayonining mohiyati, ma’lumotlarni tayyorlash, tozalash, tahlil qilish, modellash va natijalarni baholash bosqichlari keng tarzda izohlanadi. Shuningdek, klassifikatsiya, klasterlash, assotsiativ qoidalar, regressiya va boshqa usullar mazmunan tahlil qilinadi. Maqola ilmiy tadqiqotchilar, axborot tizimlari mutaxassislari hamda data mining sohasiga qiziqqan o‘quvchilar uchun mo‘ljallangan.*

Kalit so‘zlar: *Data mining, Katta ma’lumotlar, Bilimlarni aniqlash Klassifikatsiya, Klasterlash, Assotsiativ qoidalar, Analitik modellash, Ma’lumotlarni tozalash, Mashina o‘rganish, Intellektual tahlil*

KIRISH

Axborot texnologiyalari hayotning barcha jabhalariga kirib borishi bilan mavjud ma’lumotlar hajmi keskin ortdi. Korporativ tizimlarda, elektron hukumat xizmatlarida, tarmoq servislarida, ilmiy izlanishlarda, ijtimoiy tarmoqlarda va hatto kundalik dasturlardan foydalanishda ham katta hajmdagi ma’lumotlar shakllanmoqda. Ushbu ma’lumotlarda yashirin qonuniyatlar, tendensiyalar va munosabatlar mavjud bo‘lib, ularni aniqlash zamonaviy raqamli rivojlanishning muhim sharti hisoblanadi. Bunday bilimlarni aniqlashda data mining texnologiyalarining o‘rni beqiyosdir.

Data mining ma’lumotlar omborlarida, katta ma’lumotlar tizimlarida, tahlil platformalarida mavjud bo‘lgan millionlab yozuvlar orasidan foydali bilimlarni topish va ularni qaror qabul qilish jarayonida qo‘llashga imkon beradi. U nafaqat statistik yondashuvni, balki mashina o‘rganish, sun’iy intellekt, optimizatsiya, vizual analitika kabi yo‘nalishlarni ham o‘z ichiga oladi. Shuning uchun bugungi kunda data mining biznes, tibbiyot, ta’lim, bank tizimi, ishlab chiqarish, xavfsizlik va ilmiy tadqiqotlar uchun ajralmas texnologiyaga aylangan.



ASOSIY QISM

Data miningning mazmuni Data mining deganda katta hajmdagi ma'lumotlar ichidan qiymatga ega bo'lgan, foydali, amaliy ahamiyatga ega bilimlarni avtomatik yoki yarim avtomatik tarzda ajratib olish tushuniladi. Bu texnologiya ko'pincha ma'lumotlardan bilimlarni aniqlash jarayonining tarkibiy qismi sifatida qaraladi. Jarayonning maqsadi ma'lumotlar orasidagi o'xshashliklar, farqlar, bog'liqliklar, dinamik o'zgarishlar va yashirin qonuniyatlarni topishdan iborat.

Data mining uch asosiy omilga asoslanadi. Birinchisi, ma'lumotlarning keskin ko'payishi. Ikkinchisi, ma'lumotlarni chuqur tahlil qilishga bo'lgan ehtiyojning ortishi. Uchinchisi, hisoblash quvvatining yuksalishi va sun'iy intellekt usullarining keng joriy etilishi. Shu omillar tufayli data mining nafaqat ilmiy soha, balki amaliy faoliyat uchun ham muhim vositaga aylandi.

Data mining jarayonining bosqichlari

Ma'lumotlarni yig'ish va integratsiya qilish bosqichi Jarayonning dastlabki qismi ma'lumotlarni turli manbalardan yig'ish, ularni yagona formatga keltirish, moslashtirish va integratsiya qilishdan iborat. Bu bosqichda ma'lumotlar korporativ tizimlardan, sensor qurilmalardan, tajribaviy kuzatuvlardan, internet sahifalaridan, ijtimoiy tarmoqlardan yoki ma'lumotlar omborlaridan olinishi mumkin. Ma'lumotlarning turli formatlarda bo'lishi, ular orasidagi nomuvofiqliklar yoki yetishmayotgan ma'lumotlar jarayonni murakkablashtiradi.

Ma'lumotlarni tozalash va tayyorlash bosqichi Ushbu jarayon data miningning eng muhim qismlaridan biri bo'lib, umumiy vaqtning katta qismi aynan shu bosqichga sarflanadi. Ma'lumotlarning tozalangan, izchil va mantiqan mos bo'lishi tahlil natijalarining ishonchliligini oshiradi. Tozalash jarayoniga xatolarni bartaraf etish, takror ma'lumotlarni olib tashlash, yetishmayotgan qiymatlarni to'ldirish, muvofiqlashtirish va normallashtirish kiradi. Yaxshi tayyorlangan ma'lumotlar keyingi algoritmlarning samaradorligini sezilarli darajada oshiradi.

Ma'lumotlarni tanlash va atributlarni qayta ishlash Tahlil uchun barcha atributlar birdek muhim bo'lavermaydi. Shuning uchun data mining jarayonida eng ahamiyatli atributlarni tanlash, ortiqcha o'zgaruvchilarni chiqarib tashlash yoki ularni yangi atributlarga aylantirish muhim hisoblanadi. Bu bosqich ma'lumotlar o'lchamini optimallashtiradi va modellarning tezroq ishlashiga yordam beradi.

Model tanlash va qurish bosqichi Ushbu bosqichda ma'lumotlarga mos keladigan analitik model tanlanadi. Modellar mashina o'rganish algoritmlari, statistik yondashuvlar, neyron tarmoqlar yoki grafik modellardan iborat



bo'lishi mumkin. Modelning tanlanishi maqsadga bog'liq bo'ladi. Masalan, klassifikatsiya, klasterlash, prediktsiya yoki assotsiativ tahlil kabi vazifalar turli yondashuvlarni talab qiladi.

Modelni o'rgatish va sinovdan o'tkazish bosqichi Tanlangan model ma'lumotlar asosida o'rgatiladi. O'rgatish jarayonida model ma'lumotlar orasidagi bog'liqliklarni aniqlaydi, natijalarni umumlashtiradi va ko'p martalik iteratsiyalar orqali optimallashtiriladi. Sinov jarayonida esa model yangi ma'lumotlar bilan tekshiriladi, xatolik darajasi baholanadi, kuchli va zaif tomonlari aniqlanadi.

Natijalarni tahlil qilish va talqin etish Data miningning yakuniy maqsadi faqat model tuzish emas, balki amaliyotda foydalanishga yaroqli natijalarni olishdir. Shu sababli hosil bo'lgan model natijalari vizuallashtiriladi, ekspertlar tomonidan talqin qilinadi, qaror qabul qilish tizimlariga integratsiya qilinadi yoki strategik rejalashtirish jarayonida qo'llanadi.

Data miing metodlari

Klassifikatsiya Klassifikatsiya usuli mavjud ma'lumotlarni oldindan belgilangan toifalarga ajratish vazifasini bajaradi. Masalan, mijozlarni faol yoki passiv guruhlariga ajratish, xatlar orasidan spamni aniqlash, kasallik bor yoki yo'qligini bashorat qilish. Klassifikatsiyada qaror daraxtlari, qoida asosidagi tizimlar, logistik modellash, k-yaqin qo'shnilar usuli va neyron tarmoqlar keng qo'llanadi. Ushbu metod aniqlik, izohlanish osonligi va samaradorligi bilan ajralib turadi.

Klasterlash Klasterlash data miningning nazoratsiz tahlil turiga kiradi. Bu usulda ma'lumotlar oldindan belgilangan yorliqlarsiz, o'zaro o'xshashliklariga qarab guruhlariga ajratiladi. Klasterlash mijozlarni segmentlarga ajratish, bozordagi mahsulotlarni tahlil qilish, biologik ma'lumotlarni guruhlash kabi sohalarida keng qo'llanadi. Mashhur algoritmlar ichida k-markazlar usuli, ierarxik klasterlash va zichlikka asoslangan algoritmlar mavjud.

Assotsiativ qoidalar Assotsiativ tahlil ma'lumotlar orasidagi bog'liqliklarni aniqlaydi. Masalan, do'konda xarid savatlari tahlili orqali qaysi mahsulotlar birga xarid qilinishini aniqlash mumkin. Apriori va FP-tree algoritmlari ushbu jarayonning asosiy vositalaridir. Assotsiativ qoidalar tavsiya tizimlarida va marketing strategiyalarida muhim o'rin tutadi.

Regressiya Regressiya o'zgaruvchilar orasidagi munosabatni matematik jihatdan ifodalash usuli hisoblanadi. U miqdoriy bashoratlar uchun qo'llanadi. Masalan, uy narxini, bozor talabini, iqtisodiy ko'rsatkichlarning kelgusi qiymatlarini prognozlash regressiya yordamida amalga oshiriladi. Oddiy



regressiya, ko'p omilli regressiya, polynomial regressiya va penalizatsiyalangan modellash keng tarqalgan.

Neyron tarmoqlar va chuqur o'rganish Zamonaviy data miningning muhim yo'nalishlaridan biri neyron tarmoqlar va chuqur o'rganish hisoblanadi. Ushbu metodlar katta hajmdagi murakkab tuzilgan ma'lumotlardan yuqori aniqlikda bilimlarni aniqlash imkonini beradi. U tasvirlarni tanib olish, nutqni qayta ishlash, tabiiy tilni tahlil qilish, prognozlash kabi murakkab vazifalarda qo'llanadi. Chuqur o'rganish ko'p qatlamli neyron tarmoqlar orqali amalga oshiriladi va ulkan hisoblash quvvatini talab qiladi.

XULOSA

Data mining zamonaviy axborot tizimlarida ma'lumotlarni tahlil qilishning eng kuchli va samarali yo'llaridan biri hisoblanadi. U katta hajmdagi ma'lumotlar ichidan amaliy ahamiyatga ega bo'lgan bilimlarni avtomatik tarzda ajratib olish imkonini beradi. Maqolada data miningning mohiyati, bosqichlari, asosiy tahliliy metodlari keng yoritildi. Bugungi raqamli iqtisodiyotda data mining texnologiyalari ilmiy izlanishlar, biznes tahlil, sanoat jarayonlarini optimallashtirish, xavfsizlikni ta'minlash, tibbiyot diagnostikasi, ta'lim jarayonini takomillashtirish kabi ko'plab sohalarda keng qo'llanmoqda. Kelajakda data mining sun'iy intellektning yanada rivojlanishi bilan birgalikda yanada murakkab, aqlli va moslashuvchan tizimlarni yaratishga xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR:

- 1.Data Mining: Concepts and Techniques. Jiawei Han, Micheline Kamber, Jian Pei. Elsevier nashriyoti.
- 2.Machine Learning. Tom Mitchell. McGraw Hill nashriyoti.
- 3.Pattern Recognition and Machine Learning. Christopher Bishop. Springer nashriyoti.
- 4.The Elements of Statistical Learning. Hastie, Tibshirani, Friedman.
- 5.Data Mining for Business Intelligence. Galit Shmueli va boshqalar.
- 6.Introduction to Data Mining. Pang-Ning Tan, Steinbach, Kumar.
- 7.Ma'lumotlar omborlari va analitik tizimlar bo'yicha O'zbekiston oliy ta'lim darsliklari.

