



DATA MINING TEXNALOGIYALARI METODLARI VA BOSQICHLARI HAMDA DATA SCIENCE JARAYONLAR

Mamasidiqova Ismigul Abdulatif qizi

Farg'ona davlat universiteti

Amaliy matematika va informatika yo'nalishi 4-bosqich talabasi

email: ismigul0298@gmail.com

Husanova Oydinoy

Farg'ona davlat universiteti

Amaliy matematika va informatika yo'nalishi 4-bosqich talabasi

Madaminova Aziza

Farg'ona davlat universiteti

Amaliy matematika va informatika yo'nalishi 4-bosqich talabasi

Tojimatov Isroil

Ilmiy rahbar:

Fargona davlat universiteti o'qituvchisi

<https://doi.org/10.5281/zenodo.7767028>

ARTICLE INFO

Qabul qilindi: 16- mart 2023 yil

Ma'qullandi: 18-mart 2023 yil

Nashr qilindi: 24-mart 2023 yil

KEY WORDS

data mining texnologiyalari, data mining metodlari, data mining bosqichlari, data science jarayonlar, data mining.

ABSTRACT

Mazkur maqola data mining texnologiyalari metodlari va bosqichlari hamda data science jarayonlarini ko'rib chiqqan holda tahlil qilingan. Bundan tashqari, data mining nima? Data mining texnologiyalari nima? Data mining nechta metodlardan iborat? Data mining bosqichlari qanday? Data science jarayonlar qanday amalga oshiriladi? Shular xususida so'z yuritamiz. [1]

Hozirgi kunda korxonalar juda ajoyib tezlikda ma'lumotlarni yig'ishmoqda. Ushbu ulkan ma'lumotlar oqimining manbalari xilma-xildir. Bu kredit karta tranzaksiyalari, umumiy foydalanish mumkin bo'lgan mijozlar ma'lumotlari, banklar va moliya institutlari ma'lumotlari, shuningdek, foydalanuvchilar o'zlarining noutbuklari, mobil telefonlari, planshetlari va ish stollarida dasturdan foydalanish va yuklab olish uchun taqdim etishlari kerak bo'lgan ma'lumotlardan kelib chiqishi mumkin. Bunday katta hajmdagi ma'lumotlarni saqlash oson emas. Shu maqsadda ko'plab relyatsion ma'lumotlar bazasi serverlari doimiy ravishda qurilmoqda. Bularning barchasini turli ma'lumotlar bazasi serverlarida saqlash uchun onlayn tranzaksiya protokoli yoki OLTP tizimlari ham ishlab chiqilmoqda. OLTP tizimlari biznesning muammosiz ishlashiga yordam berishda muhim rol o'ynaydi. Aynan shu tizimlar ma'lumotlar bazasiga eng kichik tranzaksiyalardan kelib chiqadigan ma'lumotlarni saqlash uchun javobgardir. Shunday qilib, sotish, sotib olish, inson kapitalini boshqarish va boshqa operatsiyalar bilan bog'liq ma'lumotlar OLTP tizimlari tomonidan ma'lumotlar bazasi serverlarida saqlanadi.

Data Mining nima?

Ma'lumotni yuklab olish – bu tendentsiyalar, naqshlar va foydali ma'lumotlarni aniqlash uchun berilgan ma'lumotlar to'plamidan ma'lumot olishda yordam beradigan jarayon.

Ma'lumotni yuklab olishdan foydalanishning maqsadi katta ma'lumotlar to'plamidan ma'lumotlarga asoslangan qarorlar qabul qilishdir. Ma'lumotlarni yuklab olish maxsus guruh muammolari bilan ishlash uchun mo'ljallangan murakkab algoritmlardan foydalanadigan statistika fanining bir tarmog'i bo'lgan bashoratli tahlil bilan birgalikda ishlaydi. Bashoratli tahlil birinchi navbatda katta hajmdagi ma'lumotlarning naqshlarini aniqlaydi, bu ma'lumotlarni qazib olish bashorat va prognozlar uchun umumlashtiradi. Ma'lumotni yuklab olish o'ziga xos maqsadga xizmat qiladi, ya'ni ma'lum bir domenga tegishli bo'lgan muammolar to'plamidagi ma'lumotlar to'plamidagi naqshlarni tanib olishdir.

Ma'lumotlarning yuklab olinishi mumkin bo'lgan turlari

1. Ma'lumotlar bazasida saqlanadigan ma'lumotlar

Ma'lumotlar bazasi ma'lumotlar bazasini boshqarish tizimi yoki DBMS deb ham ataladi. Har bir ma'lumotlar bazasi ma'lumotlar bazasi bir-biri bilan bog'liq bo'lgan ma'lumotlarni saqlaydi. Shuningdek, u ma'lumotlarni boshqarish va ularga oson kirishni ta'minlash uchun foydalaniladigan dasturiy ta'minot dasturlari to'plamiga ega. Ushbu dasturiy ta'minot dasturlari juda ko'p maqsadlarga xizmat qiladi, jumladan, ma'lumotlar bazasi tuzilmasini aniqlash, saqlangan ma'lumotlarning xavfsiz va izchil bo'lishiga ishonch hosil qilish va umumiy, taqsimlangan va bir vaqtning o'zida ma'lumotlarga kirishning turli turlarini boshqarish. Relyatsion ma'lumotlar bazasida turli nomlar, atributlarga ega bo'lgan va katta ma'lumotlar to'plamining qatorlari yoki yozuvlarini saqlashi mumkin bo'lgan jadvallar mavjud. Jadvalda saqlangan har bir yozuv o'ziga xos kalitga ega. Obyektlar va ular o'rtasidagi munosabatlarni aks ettiruvchi relyatsion ma'lumotlar bazasini taqdim etish uchun ob'ekt-munosabatlar modeli yaratilgan.

2. Ma'lumotlar ombori

Ma'lumotlar ombori - bu bir nechta manbalardan ma'lumotlarni to'playdigan va keyin uni yagona reja shaklida saqlaydigan yagona ma'lumotlarni saqlash joyi. Ma'lumotlar omborida saqlangan ma'lumotlar tozalash, integratsiya, yuklash va yangilashdan o'tadi. Ma'lumotlar omborida saqlanadigan ma'lumotlar bir necha qismlarga bo'linadi. Agar siz 6 yoki 12 oy oldin saqlangan ma'lumotlar haqida ma'lumot olishni istasangiz, uni xulosa shaklida olasiz.

3. Transaksiyon ma'lumotlar

Tranzaksiya ma'lumotlar bazasi tranzaksiyalar sifatida olingan yozuvlarni saqlaydi. Ushbu tranzaksiyalarga parvozlarni bron qilish, mijoz sotib olish, veb-saytga bosish va boshqalar kiradi. Har bir tranzaksiya yozuvi noyob identifikatorga ega. Shuningdek, u tranzaksiyani amalga oshirgan barcha elementlarni sanab o'tadi.

4. Boshqa turdagi ma'lumotlar

Bizda tuzilishi, semantik ma'nolari va ko'p qirraliligi bilan mashhur bo'lgan boshqa ko'plab ma'lumotlar turlari mavjud. Ular ko'plab dasturlarda qo'llaniladi. Mana shu ma'lumotlar turlaridan bir nechtasi: ma'lumotlar oqimlari, muhandislik dizayn ma'lumotlari, ketma-ketlik ma'lumotlari, grafik ma'lumotlar, fazoviy ma'lumotlar, multimedia ma'lumotlari va boshqalar.

Ma'lumotlarni yuklab olish texnikasi

1. Assosiatiya

Bu boshqalardan eng ko'p qo'llaniladigan ma'lumotlarni qidirish usullaridan biridir. Ushbu texnikada tranzaksiya va uning elementlari o'rtasidagi munosabatlar naqshni aniqlash uchun ishlatiladi. Shuning uchun bu texnika aloqa texnikasi deb ham ataladi. U bozor savatchasini tahlil qilish uchun ishlatiladi, bu mijozlar muntazam ravishda birgalikda sotib oladigan barcha

mahsulotlarni aniqlash uchun amalga oshiriladi. Ushbu usul turli mijozlarning xarid qilish odatlarini o'rganish uchun foydalanishi mumkin bo'lgan chakana sotuvchilar uchun juda foydali. Chakana sotuvchilar o'tmishdagi savdo ma'lumotlarini o'rganishlari va keyin mijozlar birgalikda sotib oladigan mahsulotlarni qidirishlari mumkin. Keyin ular mijozlarga vaqtlarini tejash va sotishni ko'paytirishga yordam berish uchun o'zlarining chakana savdo do'konlarida ushbu mahsulotlarni bir-biriga yaqin joylashtirishlari mumkin.

2. Klasterlash

Ma'lumotlar qazib olishning yana bir metodologiyasi klasterlashdir. Bu bir xil xususiyatlarga ega bo'lgan mazmunli ob'ekt klasterlarini yaratadi. Odamlar ko'pincha uni tasniflash bilan chalkashtirib yuborishadi, lekin agar ular ushbu ma'lumotlarni qidirish metodologiyasi yoki usullari qanday ishlashini to'g'ri tushunsalar, ularda hech qanday muammo bo'lmaydi. Ob'ektlarni oldindan belgilangan sinflarga joylashtiradigan tasniflashdan farqli o'laroq, klasterlash ob'ektlarni o'zi tomonidan belgilangan sinflarga joylashtiradi. Keling, bir misol keltiraylik. Kutubxona turli mavzulardagi kitoblar bilan to'la. Endi muammo o'sha kitoblarni o'quvchilarga ma'lum bir mavzu bo'yicha kitoblarni topishda hech qanday muammo tug'dirmaydigan tarzda tartibga solishdir. Biz o'xshashliklari bo'lgan kitoblarni bitta javonda saqlash va keyin bu javonlarga mazmunli nom berish uchun klasterlashdan foydalanishimiz mumkin. Muayyan mavzu bo'yicha kitob izlayotgan o'quvchilar to'g'ridan-to'g'ri o'sha javonga borishlari mumkin. Ulardan kitoblarini topish uchun butun kutubxonani kezish talab qilinmaydi. Klaster tahlili bir-biriga o'xshash ma'lumotlarni aniqlaydi. Bu ma'lumotlar o'rtasidagi o'xshashlik va farqlarni aniqlaydi. U segmentatsiya sifatida tanilgan va ma'lumotlar bazasida sodir bo'layotgan voqealar haqida tushuncha beradi.

3. Tasniflash

Ushbu uslub o'zining kelib chiqishini mashinani o'rganishda topadi. U ma'lumotlar to'plamidagi elementlar yoki o'zgaruvchilarni oldindan belgilangan guruhlar yoki sinflarga tasniflaydi. U boshqa usullar qatorida ma'lumotlarni qazib olishda chiziqli dasturlash, statistika, qarorlar daraxtlari va sun'iy neyron tarmog'idan foydalanadi. Tasniflash ma'lumotlar to'plamidagi narsalarni turli sinflarga tasniflash qobiliyatiga ega bo'ladigan tarzda modellashtirilishi mumkin bo'lgan dasturiy ta'minotni ishlab chiqish uchun ishlatiladi. Masalan, biz undan suhbatda qatnashgan barcha nomzodlarni ikki guruhga ajratish uchun foydalanishimiz mumkin - birinchi guruh tanlangan nomzodlar ro'yxati, ikkinchisi esa rad etilgan nomzodlar ro'yxati. Ushbu tasniflash ishini bajarish uchun ma'lumotlarni qidirish dasturidan foydalanish mumkin.

5. Ketma-ket naqshlar

Ushbu uslub tranzaksiya ma'lumotlaridan foydalanishga, so'ngra ma'lum vaqt davomida undagi o'xshash tendentsiyalarni, naqshlarni va voqealarni aniqlashga qaratilgan. Tarixiy savdo ma'lumotlari xaridorlar yilning turli vaqtlarida birgalikda sotib olgan narsalarni aniqlash uchun ishlatilishi mumkin. Tarixiy ma'lumotlar shuni ko'rsatmasa ham, mijozlarga ushbu mahsulotlarni sotib olishni tavsiya qilish orqali biznes ushbu ma'lumotni tushunishi mumkin. Korxonalar ushbu tavsiyani bajarish uchun foydali bitimlar va chegirmalardan foydalanishlari mumkin.

Data science nima?

Booz Allen Hamiltonning (BAH) ma'lumotlar faniga oid qo'llanmasi bu savolga mukammal javob berishga harakat qiladi: "Ma'lumotlar fanini tasvirlash quyosh botishini tasvirlashga

o'xshaydi - bu oson bo'lishi kerak, ammo so'zlarni qandaydir tarzda yozib olish mumkin emas."Ba'zilar ma'lumot olimi Silikon vodiysidagi ma'lumotlar tahlilchisi desa, boshqalar buni statistikdan ko'ra ko'proq dasturlashni biladigan va dasturiy ta'minot ishlab chiqaruvchisidan ko'ra ko'proq statistik biladigan odam deyishadi. Boshqalar buni mashhur Venn diagrammasining o'zgarishi bilan tasvirlashgan. **Data science**- bu ma'lumotlar haqidagi fan bo'lib, sohalararo ilmiy usulublar, jarayonlar, algoritmlar va tizimlardan tashkil topgan tarkiblangan va tarkiblanmagan bilim va g'oyalarni o'z ichiga oluvchi fandır. **Data science**-ma'lumotlarni intellectual tahlil qilish, mashinani o'rganish va katta ma'lumotlar (Big data) bilan bog'liqdir. Data science - bu ma'lumotlar yordamida "haqiqiy dunyo hodisalarini tushunish va tahlil qilish" maqsadida "Statistika, ma'lumotlarni tahlil qilish va tegishli texnikalarni birlashtirish" konsepsiyalaridir. U matematika, statistika, informatika va informatika fan bilimlari kontekstida ko'plab sohalardan olingan usullar va nazariyalardan foydalanadi. Turing mukofoti sovrindori Jim Grey Data science fani "to'rtinchi paradigma" (empirik, nazariy, hisoblash va hozirda ma'lumotlarga asoslangan) fan sifatida taqdim etdiladi hamda "barcha fanlardagi hamma narsalar axborot texnologiyalari ta'siri tufayli o'zgarib boradi" va ma'lumotlar oqimining ko'payishini o'rganadi – deb ta'kidlaydi. [2]

XULOSA

Xulosa o'rnida shuni aytish joizki, biz yuqorida data mining haqida, data mining texnologiyalari haqida, data mining bosqichlaru hamda uning mezonlari haqida so'z yuritdik. Hamda data science jarayonlar nima ekanligini, va data mining va data science haqida olimlarning fikrini ham keltirib o'tdik. [3]

Foydalanilgan manbalar:

1. <https://wiki.loginom.ru/articles/data-mining.html>
2. <https://iot.ru/wiki/data-mining>
3. <https://trends.rbc.ru/trends/industry/61b359739a7947c7376ef7ce>
4. <https://www.sites.google.com/site/upravlenieznaniami/tehnologii-upravleniaznaniami/data-mining>
5. <https://habr.com/ru/post/95209/>
6. <https://wiki.loginom.ru/articles/data-mining.html>