

Zamonaviy dunyoda data miningning tutgan o'рни

Ma'rufjon Murodjon o'g'li Rixsibayev

mrixsibayev@gmail.com

Muzaffarjon Mashrabjon o'g'li Abdulaxatov

muzaffar.abdulakhatov@gmail.com

Toshkent davlat iqtisodiyot universiteti

Annotatsiya: Ushbu maqolada zamonaviy davrda data miningning roli qanchalik ortib borayotganini keltirib o'tilgan. Shuningdek, ma'lumotlar turlari va ularga ishlov berish jarayonlari ham keng yoritilgan.

Kalit so'zlar: data mining, ma'lumot, ma'lumot turlari, Apache Spark, ma'lumotlar tahlili, neyron tarmoqlar, model

The role of data mining in the modern world

Marufjon Murodjon o'g'li Rixsibayev

mrixsibayev@gmail.com

Muzaffarjon Mashrabjan o'g'li Abdulakhatov

muzaffar.abdulakhatov@gmail.com

Tashkent State University of Economics

Abstract: This article shows how the role of data mining is increasing in modern times. Also, the types of data and their processing processes are widely covered.

Keywords: data mining, data, data types, Apache Spark, data analysis, neural networks, model

KIRISH

Data mining (Ma'lumotlarni qazib olish) shuningdek, ma'lumotlarda bilim kashfiyoti (KDD) sifatida ham tanilgan, ishlov berilgan katta ma'lumotlar to'plamlaridan va boshqa qimmatli ma'lumotlarni ochish jarayonidir. Ma'lumotlarni saqlash texnologiyasining evolyutsiyasi va katta ma'lumotlarning o'sishini hisobga olgan holda, so'nggi ikki o'n yillikda ma'lumotlarni qazib olish texnikasini qo'llash shiddat bilan tezlashdi va kompaniyalarga o'zlarining xom ma'lumotlarini foydali bilimga aylantirish orqali yordam berdi. Biroq, ushbu texnologiya keng miqyosda ma'lumotlarni qayta ishlash uchun doimiy ravishda rivojlanib borayotganiga qaramay, rahbarlar hali ham kengayish va avtomatlashtirish bilan bog'liq muammolarga duch kelishmoqda.

Data mining ma'lumotlarni chuqur tahlil qilish orqali tashkilot qarorlarini qabul qilishni yaxshiladi. Ushbu tahlillarga asos bo'lgan ma'lumotlarni qidirish usullarini ikkita asosiy maqsadga bo'lish mumkin; ular maqsadli ma'lumotlar to'plamini tavsiflashlari yoki mashinani o'rganish algoritmlari yordamida natijalarni bashorat qilishlari mumkin. Ushbu usullar ma'lumotlarni tartibga solish va filtrlash, firibgarlikni aniqlashdan tortib foydalanuvchi xatti-harakatlari, to'siqlar va hatto xavfsizlik buzilishigacha bo'lgan eng qiziqarli ma'lumotlarni ko'rsatish uchun ishlatiladi.

Apache Spark kabi ma'lumotlarni tahlil qilish va vizualizatsiya vositalari bilan birgalikda ma'lumotlarni qazib olish dunyosiga kirish hech qachon oson bo'lmagan va tegishli tushunchalarni olish hech qachon tezroq bo'lmagan. Sun'iy intellekt sohasidagi yutuqlar faqat sanoat bo'ylab qabul qilishni tezlashtirmoqda.

Data Mining jarayoni. Data mining jarayoni katta ma'lumotlar to'plamlaridan qimmatli ma'lumotlarni olish uchun ma'lumotlarni yig'ishdan tortib vizualizatsiyagacha bo'lgan bir qancha bosqichlarni o'z ichiga oladi. Yuqorida ta'kidlab o'tilganidek, ma'lumotlarni qidirish usullari maqsadli ma'lumotlar to'plami haqida tavsif va bashoratlarni yaratish uchun ishlatiladi. Ma'lumotlar olimlari ma'lumotlarni naqshlar, assotsiatsiyalar va korrelyatsiyalarni kuzatish orqali tasvirlaydilar. Ular, shuningdek, tasniflash va regressiya usullari orqali ma'lumotlarni tasniflaydi va klasterlaydi, shuningdek, spamni aniqlash kabi foydalanish holatlari uchun o'zgarmaydiganlarni aniqlaydi.

Ma'lumotlarni qidirish odatda to'rtta asosiy bosqichdan iborat: maqsadlarni belgilash, ma'lumotlarni yig'ish va tayyorlash, ma'lumotlarni qidirish algoritmlarini qo'llash va natijalarni baholash.

1. *Biznes maqsadlarini belgilang:* Bu ma'lumotlarni qidirish jarayonining eng qiyin qismi bo'lishi mumkin va ko'plab tashkilotlar ushbu muhim bosqichga juda oz vaqt sarflaydi. Ma'lumotlar bo'yicha olimlar va biznes manfaatdor tomonlar biznes muammosini aniqlash uchun birgalikda ishlashi kerak, bu ma'lum bir loyiha uchun ma'lumotlar savollari va parametrlarini xabardor qilishga yordam beradi. Tahlilchilar, shuningdek, biznes kontekstini to'g'ri tushunish uchun qo'shimcha tadqiqotlar o'tkazishlari kerak bo'lishi mumkin.

2. *Ma'lumotlarni tayyorlash:* Muammo doirasi aniqlangandan so'ng, ma'lumotlar bo'yicha olimlar uchun qaysi ma'lumotlar to'plami biznesga tegishli savollarga javob berishini aniqlash osonroq bo'ladi. Tegishli ma'lumotlarni to'plagandan so'ng, ma'lumotlar dublikatlar, etishmayotgan qiymatlar va chegaralar kabi shovqinlarni olib tashlagan holda tozalanadi. Ma'lumotlar to'plamiga qarab, o'lchamlar sonini kamaytirish uchun qo'shimcha qadam qo'yilishi mumkin, chunki juda ko'p xususiyatlar har qanday keyingi hisoblashni sekinlashtirishi mumkin. Ma'lumotlar bo'yicha olimlar har qanday modelda optimal aniqlikni ta'minlash uchun eng muhim bashorat qiluvchilarni saqlab qolishga intiladi.

3. *Modelni yaratish va namunani qazib olish:* Tahlil turiga qarab, ma'lumotlar olimlari ketma-ket naqshlar, assotsiatsiya qoidalari yoki korrelyatsiya kabi har qanday qiziqarli ma'lumotlar munosabatlarini tekshirishlari mumkin. Yuqori chastotali naqshlar kengroq ilovalarga ega bo'lsa-da, ba'zida ma'lumotlardagi og'ishlar potentsial firibgarlik sohasini ta'kidlab, yanada qiziqarli bo'lishi mumkin.

Mavjud ma'lumotlarga qarab ma'lumotlar to'plamini tasniflash yoki klasterlash uchun chuqur o'rganish algoritmlari ham qo'llanilishi mumkin. Agar kirish ma'lumotlari etiketlangan bo'lsa (ya'ni nazorat ostida o'rganish), ma'lumotlarni toifalarga ajratish uchun tasniflash modelidan foydalanish mumkin yoki muqobil ravishda ma'lum bir topshiriqning ehtimolini bashorat qilish uchun regressiya qo'llanilishi mumkin. Agar ma'lumotlar to'plami belgilanmagan bo'lsa (ya'ni, nazoratsiz o'rganish), o'quv majmuasidagi alohida ma'lumotlar nuqtalari asosiy o'xshashliklarni aniqlash uchun bir-biri bilan taqqoslanadi va ularni shu xususiyatlar asosida klasterlanadi.

4. *Natijalarni baholash va bilimlarni amalga oshirish:* Ma'lumotlar jamlangandan so'ng, natijalarni baholash va sharhlash kerak. Natijalarni yakunlashda ular haqiqiy, yangi, foydali va tushunarli bo'lishi kerak. Ushbu mezonlar bajarilganda, tashkilotlar ushbu bilimlardan o'z maqsadlariga erishish uchun yangi strategiyalarni amalga oshirish uchun foydalanishlari mumkin.



1-rasm. Ma'lumotlarni qazib olishning to'rt bosqichi

Data mining texnikasi. Ma'lumotlarni qazib olish katta hajmdagi ma'lumotlarni foydali ma'lumotlarga aylantirish uchun turli xil algoritmlar va usullardan foydalangan holda ishlaydi. Bu erda eng keng tarqalgan ba'zilari:

Assotsiatsiya qoidalari: Assotsiatsiya qoidasi ma'lum ma'lumotlar to'plamidagi o'zgaruvchilar orasidagi munosabatlarni topish uchun qoidalarga asoslangan usuldir. Ushbu usullar tez-tez bozor savatini tahlil qilish uchun ishlatiladi, bu kompaniyalarga turli xil mahsulotlar o'rtasidagi munosabatlarni yaxshiroq tushunishga imkon beradi. Mijozlarning iste'mol odatlarini tushunish korxonalariga yanada yaxshi o'zaro sotish strategiyalari va tavsiya mexanizmlarini ishlab chiqish imkonini beradi.

Neyron tarmoqlar: Asosan chuqur o'rganish algoritmlari uchun foydalanilgan neyron tarmoqlar inson miyasining tugun qatlamlari orqali o'zaro bog'liqligini taqlid qilish orqali o'quv ma'lumotlarini qayta ishlaydi. Har bir tugun kirishlar, og'irliklar, egilish (yoki chegara) va chiqishdan iborat. Agar ushbu chiqish qiymati belgilangan chegaradan oshsa, u tarmoqdagi keyingi qatlama ma'lumotlarni uzatib, tugunni "yondiradi" yoki faollashtiradi. Neyron tarmoqlar ushbu xaritalash funksiyasini nazorat ostida o'rganish orqali o'rganadi, gradient tushish jarayoni orqali yo'qotish

funktsiyasiga asoslanadi. Xarajat funktsiyasi nolga yaqin yoki nolga yaqin bo'lsa, biz to'g'ri javob berish uchun modelning to'g'riligiga amin bo'lishimiz mumkin.

Qarorlar daraxti: Ushbu ma'lumotni qazib olish texnikasi qarorlar to'plamiga asoslangan potentsial natijalarni tasniflash yoki bashorat qilish uchun tasniflash yoki regressiya usullaridan foydalanadi. Nomidan ko'rinib turibdiki, u ushbu qarorlarning potentsial natijalarini aks ettirish uchun daraxtga o'xshash vizualizatsiyadan foydalanadi.

K-eng yaqin qo'shni (KNN): K-eng yaqin qo'shni, KNN algoritmi sifatida ham tanilgan, parametrik bo'lmagan algoritm bo'lib, ma'lumotlar nuqtalarini ularning yaqinligi va boshqa mavjud ma'lumotlar bilan bog'lanishiga qarab tasniflaydi. Ushbu algoritm o'xshash ma'lumotlar nuqtalarini bir-biriga yaqin joyda topish mumkinligini taxmin qiladi. Natijada, u odatda Evklid masofasi orqali ma'lumotlar nuqtalari orasidagi masofani hisoblashga intiladi va keyin u eng tez-tez uchraydigan toifa yoki o'rtacha toifaga asoslangan toifani tayinlaydi.

Foydalanilgan adabiyotlar

1. "Data Mining Concepts and Techniques" Third Edition by Jiawei Han, Micheline Kamber Jian Pei.
2. Fundamentals of Business Intelligence (Data-Centric Systems and Applications) 2015th Edition by Wilfried Grossmann (Author), Stefanie Rinderle-Ma (Author)..
3. "Business Intelligence – Grundlagen und praktische Anwendungen: Eine Einführung in die IT" by Hans-Georg Kemper and Henning Baars.
4. David Loshin Morgan, Kaufman, "Business Intelligence: The Savvy Manager"s Guide", Second Edition, 2012 4. Sharda, R., Delen, D. y Turban, E. (2014). Business Intelligence, A Managerial Perspective on Analytics. Boston: Pearson.
5. Bekmuradov A.Sh., Musaliyev A.A., Xashimxodjayev Sh.I. "Axborot biznesi": O'quv qo'llanma. – T.: «IQTISODIYOT», 2019. – 160 b.