O‘ZBEKISTON RESPUBLIKASI OLIY VA O‘RTA MAXSUS TA’LIM VAZIRLIGI

BUXORO DAVLAT UNIVERSITETI

F.X. Xazratov, J.J. Atamuradov, H.I. Eshonqulov

BIG DATA VA MA`LUMOTLAR TAHLILI

O`quv qo`llanma Axborot tizimlarining matematik va dasturiy ta’minoti yo‘nalishi uchun mo‘ljallangan.

Buxoro 2021

ANNOTATSIYA

F.X. Xazratov, J.J. Atamuradov, H.I. Eshonqulov Big data va ma`lumotlar tahlili. O‘quv qo‘llanma – Buxoro, BuxDU, 2021, 160 b.

Mazkur o‘quv qo‘llanma bakalavriatura 5330100 Axborot tizimlarining matematik va dasturiy ta’minoti yo‘nalishi uchun O‘zbekiston Respublikasi oliy va o‘rta maxsus ta’lim vazirligi 2018 yil “25” avgustdagi 744 -sonli buyrug‘i bilan tasdiqlangan “Big data va ma`lumotlar tahlili” namunaviy fan dasturiga muvofiq tuzilgan.

Ushbu o‘quv qo‘llanmada talabalarga davlat ta’lim standartlariga mos bilim va ko‘nikmalarni hosil qilishni ta’minlaydi, dunyoqarash va tizimli fikrlashni shakllantirishga ko‘maklashadi, hamda shu sohadagi mutaxassislarga zamonaviy dasturlash texnologiyalari usullarini va talablarga javob beradigan yuqori sifatli dasturiy ta’minotni o‘rgatadi. O‘quv qo‘llanma 3-bosqich Axborot tizimlarining matematik va dasturiy ta’minoti yo‘nalishining talabalari uchun mo‘ljallangan bo‘lib unda dasturiy tizimlarni loyihalashdagi ob’ektli yondashishning konseptsiyalari, dasturiy injenering usullari, ob’ektga yo‘naltirilgan tillar yordamida tizimli loyihalash usullarini amaliyotda tadbiq etish, korporativ ilovalarini ishlab chiqish, dasturiy komplekslarini loyihalashda kollektiv ishlab chiqish usullarini, modellashtirish tillaridan foydalangan holda predmet sohalarini tahlil etishni, diagrammallarni qurish va tizimli tahlil hamda loyihalash usullarini bilishi va ulardan foydalana olishi, dasturiy tizimlarni tizimli tahlil va loyihalashning asosiy prinsiplari va qoidalari; dasturiy muhitlarni ishlab chiqishda zamonaviy ob’ektli modellashtirishning vositalari, UML ni tadbiq qilish bo‘yicha nazariy ma‘lumotlar berilgan.

Taqrizchilar:

O.I. Jalolov - Buxoro davlat universiteti, f-m.f.n dotsent.

Sh.S. Yo`ldoshev- Buxoro muhandislik texnologiyaliri istituti, f-m.f.n dotsent.

АННОТАЦИЯ

Ф.Х. Хазратов, Ж.Ж. Атамурадов, Х. И. Эшонкулов Big data и анализ данных. Учебное пособие - Бухара, Бухарский государственный университет, 2021 160 стр.

Учебник утвержден приказом Министерства высшего и среднего специального образования Республики Узбекистан от 25 августа 2018 года № 744 на степень бакалавра 5330100 Математическое и программное обеспечение информационных систем модельной научной программы «Big data и анализ данных».

Это пособие дает студентам знания и навыки, необходимые для соответствия государственным образовательным стандартам, помогает им формировать мировоззрение и системное мышление, а также предоставляет специалистам в этой области современные технологии программирования и отвечающее требованиям высококачественное программное обеспечение.

Учебник предназначен для студентов факультетов Математическое и программное обеспечение информационных систем 3 курса и теоретическую информацию по концепции объектно-ориентированного подхода к проектированию программных систем, методы программной инженерии, методы системного проектирования с использованием объектно-ориентированных языков, изучить и использовать методы реализации, разработки корпоративных приложений, методы коллективной разработки при проектировании программных пакетов, анализ предметных областей с использованием языков моделирования, построение и систематический анализ и проектирование диаграмм, основные принципы и правила системного анализа и проектирования программных систем; средства современного объектного моделирования при разработке программных сред, внедрение UML.

Рецензенты:

О. Жалолов - Бухарский государственный университет, к.ф-м.н доцент.

Ш.С. Юлдашев - Бухарский инженерно-технологический институт, к.ф-м.н доцент.

ANNOTATION

F.Kh. Khazratov, J.J. Atamuradov, Kh. I. Eshonkulov Big data and data analysis. Textbook - Bukhara, Bukhara State University, 2021 160 pp.

The textbook was approved by order of the Ministry of Higher and Secondary Specialized Education of the Republic of Uzbekistan dated August 25, 2018 No. 744 for a bachelor's degree 5330100 Mathematical and software for information systems of the model scientific program "Big data and data analysis".

This manual provides students with the knowledge and skills necessary to meet government educational standards, helps them shape their worldview and systems thinking, and provides professionals in this field with modern programming technologies and high quality software that meets the requirements. The textbook is intended for students of the faculties of Mathematical and software of information systems of the 3rd year and theoretical information on the concept of an object-oriented approach to the design of software systems, methods of software engineering, methods of system design using object-oriented languages, to study and use methods of implementation, development of corporate applications , methods of collective development in the design of software packages, analysis of subject areas using modeling languages, construction and systematic analysis and design of diagrams, basic principles and rules of system analysis and design of software systems; means of modern object modeling in the development of software environments, the implementation of UML.

Reviewers:

O.I. Jalolov - Bukhara State University, Candidate of Physical and Mathematical Sciences, Associate Professor.

Sh.S. Yuldashev - Bukhara Institute of Engineering Technology, Candidate of Physical and Mathematical Sciences, Associate Professor.

MUNDARIJA

ОГЛАВЛЕНИЕ

TABLE OF CONTENTS

KIRISH

Ushbu o`quv qo`llanma talabalarni Data Mining texnologiyasi bilan tanishtiradi, Data Mining-ning usullari, vositalari va qo'llanilishini batafsil o'rganadi. Har bir usulning tavsifi uni qo'llashning aniq namunasi bilan birga keladi.

Data Mining va klassik statistik tahlil usullari va OLAP tizimlari o'rtasidagi farqlar muhokama qilinadi va Data Mining tomonidan aniqlangan naqsh turlari (assotsiatsiya, tasniflash, ketma-ketlik, klasterlash, prognozlash) ko'rib chiqiladi. Data Mining ilovasining doirasi tavsiflangan. Web Mining tushunchasi joriy etilgan. Ma'lumotlarni qidirish usullari batafsil ko'rib chiqiladi: neyron tarmoqlari, qarorlar daraxtlari, chegaralangan hisoblash usullari, genetik algoritmlar, evolyutsion dasturlash, klaster modellari, estrodiol usullar. Har bir usul bilan tanishish Data Mining texnologiyasidan foydalangan holda vosita yordamida amaliy vazifani hal qilish orqali tasvirlanadi. Ma'lumotlar omborlari va ularning arxitekturasidagi Data Miningning asosiy tushunchalari bayon qilingan. OLTP, OLAP, ROLAP, MOLAP tushunchalari kiritilgan va Data Mining texnologiyasidan foydalangan holda ma'lumotlarni tahlil qilish jarayoni muhokama qilinadi. Ushbu jarayonning bosqichlari batafsil muhokama qilinadi. Analitik dasturiy ta'minot bozori tahlil qilinadi, etakchi Data Mining ishlab chiqaruvchilarining mahsulotlari tavsiflanadi va ularning imkoniyatlari muhokama qilinadi.

1-MAVZU
“BIG DATA VA MA`LUMOTLAR TAHLILI” FANINING MAZMUNI, PREDMETI VA METODI

Reja:

1. "Big data va ma`lumotlar tahlili" fanining mazmuni

2. Big Data bilan ishlash bosqichlari

3. Data Mining. Data Mining texnologiyalari

**Mashg`ulot maqsadi:** *Mashg`ulotda Data Mining kontseptsiyasi batafsil muhokama qilinadi. Data Mining ning kelib chiqishi, istiqbollari, muammolari tasvirlangan. Axborot texnologiyalari bozorining bir qismi sifatida Data Mining texnologiyasiga qarash berilgan.*

Tayanch iboralar: *ma'lumotlar, Data Mining, Ma'lumotlar, tahlil, naqsh, bilimlarni ekstraksiya qilish, naqsh, bilimlarni kashf etish, KDD, statistika, namunalarni tan olish, sun'iy intellekt, sun'iy intellekt, DBMS, IMS, IBM, konferentsiya, ma'lumotlar tizimi, tarmoq modeli, SQL, interfeys, qaror qabul qilish, ma'no, bilim, ma'lumotni joylashtirish, ta'rifi, SAS, guruh, qidiruv, taqdimot, Business Intelligence, DSS, qarorlarni qo'llab-quvvatlash tizimi, axborot-tahlil tizimi, EIS, korxona ma'lumotlari, bog`liqlik.*

1. "Big data va ma`lumotlar tahlili" fanining mazmuni

Big data(katta ma'lumotlar) - juda katta hajmdagi bir jinsli bo'lmagan va tez tushadigan raqamli ma'lumotlar bo'lib, ularni odatiy usullar bilan qayta ishlab bo'lmaydi. Ba'zi hollarda, katta ma'lumotlar tushunchasi bilan birga shu ma'lumotlarni qayta ishlash ham tushuniladi. Asosan, analiz obyekti katta ma'lumotlar deb ataladi. 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. Katta ma'lumotlar (Big data) - bu bitta kontekstdagi doimiy ravishda o'sib boradigan ma'lumotlar hajmining, ammo taqdimotning turli formatlari, shuningdek, tezkor qayta ishlash usullari va vositalari. Katta ma'lumotlar: qaysi ma'lumotlar katta deb hisoblanadi. Mur qonunida tasvirlangan hisoblash quvvatining eksponentli o'sishi sababli, ma'lumotlar miqdori ularning katta yoki yo'qligini aniq mezon bo'lishi mumkin emas. Masalan, bugungi kunda katta ma'lumotlar terabaytlarda, ertaga petabaytlarda o'lchanadi. Shuning uchun Big Data-ning asosiy xususiyati bu ularning tuzilish darajasi va taqdimot variantlari.

Shakl: 1.1.1. Big dataning asosiy xususiyatlari

Sensorlardan yoki audio va video yozuv qurilmalaridan doimiy ravishda keladigan ma'lumotlar, ijtimoiy tarmoqlardan kelgan xabarlar oqimlari, meteorologik ma'lumotlar, uyali aloqa abonentlarining geolokatsion koordinatlari va boshqalar kata hajmdagi ma'lumotlarning yorqin misolidir. Masalan, bu yerda "Gazpromneft" neft quduqlaridagi boshqaruv tizimlari nazoratchilarining 200 milliondan ortiq turli xil yozuvlarini, avariya holatlaridagi kuchlanishni qayta tiklash yozuvlarini, nasos ishining o'ziga xos xususiyatlarini va nosozliklar sabablari to'g'risida farazlarni shakllantirish va sinash uchun quduq sharoitlarining

xususiyatlarini qanday to'plashi va tahlil qilishi hamda nasos uskunalarini ishlatishda ilgari noma'lum munosabatlarni aniqlash kabi vazifalarni o`z ichiga oladi. Katta ma'lumotlar manbalari quyidagicha bo'ladi: Internetdagi ijtimoiy tarmoqlar, bloglar, OAV, forumlar, veb-saytlar, (Internet of Things (IoT)); korporativ ma'lumotlar - bitimlar, arxivlar, ma'lumotlar bazalari va fayllarni saqlash; asboblarning ko`rsatgichlari - sensorlar, magnitafonlar va boshqalar.

2. Big Data bilan ishlash bosqichlari

Muayyan vaziyatlarning sabablari, xususan, uskunaning ishlamay qolishi kuchlanish sharoitlari bilan bog'liq ishchi farazni olish yoki kelajakni bashorat qilish uchun, masalan, xususiy qarz oluvchi tomonidan qarzni o'z vaqtida qaytarish ehtimoli, tuzilgan va tuzilmagan ma'lumotlarning katta hajmini tahlil qilish bir necha bosqichlarda amalga oshiriladi.

1. ma'lumotlarni tozalash - ma'lumotlarning dastlabki to'plamidagi xatolarni qidirish va tuzatish, masalan, qo'lda kiritish xatolari, qisqa muddatli nosozliklar tufayli o'lchash moslamalarining noto'g'ri qiymatlari va hk.;
2. bashorat qiluvchilar avlodi (xususiyat muhandisligi) - analitik modellarni qurish uchun o'zgaruvchilar, masalan, ma'lumot, potentsial qarz oluvchining jinsi va yoshi;
3. maqsad o'zgaruvchini bashorat qilish uchun analitik modelni (modelni tanlash) qurish va o'rgatish. Shunday qilib, maqsad o'zgaruvchisining predikatorlarga bog'liqligi haqidagi farazlar qanday tekshiriladi. Masalan, o'rta ma'lumotli va 3 oydan kam ish tajribasiga ega bo'lgan qarz oluvchi uchun qarzni to'lash muddati necha kun.

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;

- krodsourcing - 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.

3. Data Mining. Data Mining texnologiyalari

Data mining(ma'lumotlarni topish) - biron qonuniyatni topish maqsadida ma'lumotlarni intellectual analiz qilishga aytiladi. Isroillik matematik Grigoriy Pyatetskiy-Shapiro 1989-yilda bu atamani fanga kiritgan.

Texnologiyalar, avvalari noma'lum va foydali bo'lgan qayta ishlanmagan(hom) ma'lumotlarni topish jarayoniga data mining(ma'lumotlarni topish) deyiladi. Data mining metodlari ma'lumotlar ombori, statistika va sun'iy intellekt tutashgan nuqtada joylashadi.

Ma'lumotlar qidirish usullari har xil tasniflash, modellashtirish va prognoz qilish usullariga asoslangan bo'lib, qaror daraxtlarini, sun'iy neyron tarmoqlarini, genetik algoritmlarni, evolyutsion dasturlashni, assotsiativ xotirani, loyqa mantiqni

ishlatishga asoslangan. Ma'lumotlarni qidirish usullari ko'pincha statistik usullarni o'z ichiga oladi (tavsifli tahlil, korrelyatsiya va regressiya tahlili, omillar tahlili, tafovutni tahlil qilish, tarkibiy qismlarni tahlil qilish, diskriminant tahlil, vaqtni tahlil qilish, yashashni tahlil qilish, havolani tahlil qilish). Ammo bunday usullar tahlil qilingan ma'lumotlar haqida ba'zi bir afsonaviy fikrlarni qabul qiladi, bu ma'lumotlar qidirish maqsadlariga (ilgari noma'lum bo'lmagan va amaliy foydali bilimlarni kashf etish) zid keladi.

Ma'lumotlar qidirish usullarining eng muhim maqsadlaridan biri bu maxsus matematik tayyorgarlikka ega bo'lmagan odamlar tomonidan ma'lumotlarni qidirish vositalaridan foydalanishga imkon beradigan hisob-kitoblarning natijalari (vizualizatsiya).

Dastlab, ma'lumotlar bazasi mavjud vazifa quyidagicha belgilanadi:

- juda katta;
- ma'lumotlar bazasida ba'zi "yashirin bilimlar" mavjud deb taxmin qilinadi.

Katta hajmdagi tayyor bo`lmagan ma'lumotlarda yashirin bo'lgan metodlarini aniqlash usullarini ishlab chiqish kerak. Hozirgi global raqobat sharoitida qo'shimcha raqobatbardosh ustunlik manbai bo'lishi mumkin bo'lgan aniqlangan ilmlar (bilimlar)ni aniq ishlab chiqish kerak bo`ladi

"Yashirin bilim" nimani anglatadi?

- ilgari noma'lum - ya'ni yangi bo'lishi kerak bo'lgan bilim (va ilgari olingan ma'lumotni tasdiqlamagan holda);
- trivial emas - bu shunchaki ko'rinmaydigan narsalar (ma'lumotlarni to'g'ridan-to'g'ri vizual tahlil qilish bilan yoki oddiy statistik tavsiflarni hisoblashda);
- amaliy jihatdan foydali - ya'ni tadqiqotchi yoki iste'molchi uchun qadrli bo'lgan bilimlar;
- sharhlash uchun qulay - bu foydalanuvchi uchun vizual ko'rinadigan va mavzu doirasi bo'yicha tushuntirishga oson bo'lgan bilim.

Ushbu talablar ma'lumotlar qidirish usullarining mohiyatini va ma'lumotlar yig'ish texnologiyasida ma'lumotlar bazasini boshqarish tizimlari, statistik tahlil

usullari va sun'iy intellekt usullaridan qaysi shaklda va qaysi nisbatda foydalanilishini aniqlaydi.

Ma'lumot olish usullari katta ma'lumotlar bilan ishlashda ham, nisbatan kam miqdordagi ma'lumotlarni qayta ishlashda ham qo'llanilishi mumkin (masalan, individual eksperimentlar natijalari natijasida yoki kompaniyaning faoliyati to'g'risidagi ma'lumotlarni tahlil qilishda olinadi) Yetarli miqdordagi ma'lumotlarning mezoni sifatida tadqiqot sohasi, va amaliy tahlil algoritmi kerak bo`ladi.

Data mining yordamida muammolarni hal qilishning bir qator bosqichlari:

1. Tahlil vazifasi bayoni;
2. Ma'lumot to'plash;
3. Ma'lumotlarni tayyorlash (filtrlash, qo'shish, kodlash);
4. Modelni tanlash (ma'lumotlarni tahlil qilish algoritmi);
5. Model parametrlari va o'rganish algoritmini tanlash;
6. Modelni o'qitish (boshqa model parametrlarini avtomatik izlash);
7. Agar 5-bandga yoki 4-bandga o'tish qoniqarli bo'lmasa, o'qitish sifatini tahlil qilish;
8. 1, 4 yoki 5-bandlarga o'tish qoniqarsiz bo'lsa, aniqlangan bilimlarni tahlil qilish.

Data mining texnik xarakteristikasi

Ma'lumotni izlash asosan uchta tushunchaga asoslanadi:

- Matematik statistika ma'lumotlarni yig'ishda ishlatiladigan texnologiyalarning asosini tashkil etadi, masalan, klasterli tahlil, regression tahlil, diskriminatsion tahlil va boshqalar.
- Sun'iy intellekt - inson fikrlaydigan neyron tarmog'ini raqamli ko'paytirish;
- Mashinalarni o'rganish - bu eng mos keladigan tahlil usulini yoki qulay usulni tanlash uchun kompyuterlarga qayta ishlanadigan ma'lumotlarni tushunishga yordam beradigan statistika va sun'iy intellektlar to'plami.

Ma'lumotlarni qidirishda quyidagi asosiy vazifalar sinflari qo'llaniladi:

- og'ishlarni aniqlash - ba'zi parametrlarda umumiy massadan farq qiladigan ma'lumotlarni aniqlash;
- uyushma mashg'ulotlari - voqealar o'rtasidagi munosabatlarni topish;
- klasterlash - oldindan ma'lum bo'lgan naqshlarsiz ma'lumotlar to'plamlarini guruhlash;
- tasniflash - yangi ma'lumotlarga murojaat qilish uchun ma'lum bilimlarni umumlashtirish;
- regressiya - ma'lumotlar to'plamini eng kichik og'ish bilan ko'rsatadigan funktsiyani topish;
- umumlashtirish - dastlabki ma'lumotlarni siqilgan shaklida ko'rsatish, shu jumladan hisobotlarni taqdim etish va vizualizatsiya.

Nazorat savollari

1. Big data deganda nimani tushunasiz?
2. Mur qanday qoniniyat yaratgan?
3. Big data ning asosiy bosqichlariga misollar keltiring.
4. Qanday big dataning to'plash va tahlil qilishning asosiy usullarini bilasiz?

2-MAVZU

MA`LUMOTLAR

Reja:

1. **Ma`lumotlar bazasi**
2. **Ma`lumotlar turlari**
3. **Ma`lumotlarni saqlash formatlari**
4. **Metama`lumotlar**

**Mashg`ulot maqsadi:** *Mashg`ulotda ma'lumotlar tushunchasi batafsil muhokama qilinadi. Ob'yekt va atribut, tanlov, qaram va mustaqil o'zgaruvchilar tushunchalarining ma'nosi tushuntiriladi. Tarozi turlari batafsil muhokama qilinadi. Turli xil ma'lumotlar to'plamlari berilgan. Ma'lumotlar bazasi va DBMS tushunchalari qisqacha ko'rib chiqildi.*

Tayanch iboralar: *ma'lumotlar, jadval, dasturiy ta'minot, atribut, ob'yekt, yozuv, namuna, populyatsiya, gipoteza, naqsh, munosabatlar, joy, o'zgaruvchi, o'lchov, ma'lumotlar, navbat, qiymat, aniqlik, balandlik, vazn, uzunlik, ma'lumotlar to'plami, ta'rif, hisoblash, fayl, matn, CSV, bo'shliq, yorliqlar bilan ajratilgan qiymatlar, SAS, Excel, bilim, qidirish, kirish, ma'lumotlar bazasi menejeri, ob'yekt paskal, mantiqiy dizayn, kaskadli o'chirish, OLAP, doimiylar, koordinatalar, foyda, qoldiq, so'rov, metama'lumotlar.*

1. Ma`lumotlar bazasi

1968 yilda IBM-dan birinchi IMS sanoat DBMS tizimi foydalanishga topshirildi.1975 yilda ma'lumotlar qayta ishlash tillari assotsiatsiyasining birinchi standarti - Ma'lumotlar tizimlari tillari bo'yicha konferentsiya (CODASYL) paydo bo'ldi, u ma'lumotlar bazasi tizimlari nazariyasida tarmoq ma'lumotlari modeli uchun hali ham muhim bo'lgan bir qator fundamental tushunchalarni aniqladi. Ma'lumotlar bazasi nazariyasini yanada rivojlantirishga amerikalik matematik E.F. Codd ma'lumotlar modelini yaratgan. Ushbu davrda ko'plab tadqiqotchilar ma'lumotlar bazalariga tuzilish va kirishni ta'minlash yo'nalishida yangi

yondashuvni sinab ko'rishdi. Ushbu qidiruvlarning maqsadi ma'lumotlarni osonroq modellashtirish uchun relyatsion prototiplarni olish edi. Natijada 1985 yilda SQL deb nomlangan til yaratildi. Bugungi kunda deyarli barcha ma'lumotlar bazalari ushbu interfeysni ta'minlaydilar.

Ma'lum ma'lumotlarning turlari paydo bo'ldi - "grafik tasvir", "hujjat", "tovush", "xarita". SQL tiliga vaqt, vaqt oralig'i, ikki baytli belgilar satrlari uchun ma'lumotlar turlari qo'shildi. DataMining texnologiyalari, ma'lumotlar omborlari, multimedia ma'lumotlar bazalari va veb-ma'lumotlar bazalari paydo bo'ldi.

Ma'lumot konining paydo bo'lishi va rivojlanishi turli omillarga bog'liq, ularning asosiylari quyidagilardan iborat:

apparat va dasturiy ta'minotni takomillashtirish;

- ma'lumotlarni saqlash va qayd etish texnologiyalarini takomillashtirish;
- katta miqdordagi tarixiy ma'lumotlarni to'plash;
- axborotni qayta ishlash algoritmlarini takomillashtirish.

Ma'lumotlar bazasi. Asosiy qoidalar

Ma'lumotlar bazasida ma'lumotlarni tashkil qilishni tushunish ma'lumotlar bazasi nazariyasining asoslarini bilishni talab qiladi. Keling, ushbu nazariyaning ba'zi qoidalarini ko'rib chiqaylik.

Ma'lumotlar bazasi (Database) - bu maxsus tashkil etilgan va elektron shaklda saqlanadigan ma'lumotlar.

Maxsus tashkil etilgan deganda ma'lumotlar bir yoki bir nechta ilovalarni topish va ulardan foydalanishni osonlashtiradigan maxsus tarzda tashkil etilganligini anglatadi. Shuningdek, bunday ma'lumotlarni tashkillashtirish ma'lumotlarning minimal zaxirasini ta'minlaydi.

Ma'lumotlar bazalari - bu axborot texnologiyalarining turlaridan biri, shuningdek ma'lumotlarni saqlash shakli. Ma'lumotlar bazasini yaratishning maqsadi dasturiy ta'minotga, foydalaniladigan texnik vositalarga va kompyuterdagi ma'lumotlarning jismoniy joylashishiga bog'liq bo'lmagan ma'lumotlar tizimini yaratishdir. Bunday ma'lumotlar tizimining qurilishi izchil va to'liq ma'lumotlarni taqdim etishi kerak. Ma'lumotlar bazasini loyihalashda undan ko'p maqsadli foydalanish qabul qilinadi.

Eng oddiy holatda ma'lumotlar bazasi ikki o'lchovli jadvallar tizimi sifatida taqdim etiladi.

Ma'lumotlar sxemasi - bu ma'lumotlar tavsifi tilida ko'rsatilgan va MBBT tomonidan ishlov berilgan mantiqiy ma'lumotlar strukturasining tavsifi.

Foydalanuvchi sxemasi - ma'lum bir foydalanuvchi uchun o'rnatiladigan jadval maydonchasi tartibining bitta varianti.

Ma'lumotlar bazasini boshqarish tizimlari, MBBT

Ma'lumotlar bazasini boshqarish tizimi - bu ma'lumotlar bazasida ma'lumotlarni tashkil qilish, saqlash, yaxlitlik, o'zgartirish, o'qish va xavfsizligini boshqaruvchi dastur.

DBMS (Database Management System) bu qobiq bo'lib, uning yordamida jadvallarning tuzilishini tashkil etish va ularni ma'lumotlar bilan to'ldirishda u yoki bu ma'lumotlar bazasi olinadi.

Relational Database Management System - bu relyatsion ma'lumotlar modeliga asoslangan ma'lumotlar bazasi. Relyatsion ma'lumotlar modelida har qanday ma'lumotlar vakili relyatsion jadvallar to'plamiga qisqartiriladi (maxsus tipdagi ikki o'lchovli jadvallar). Ma'lumotlar omborini yaratish uchun ma'lumotlar bazasini boshqarish tizimlari qo'llaniladi. Ma'lumotlar bazasi dasturiy, texnik va tashkiliy qismlarga ega. Dasturiy ta'minot ma'lumotlar bazasini kiritish-chiqarish, ma'lumotlarni qayta ishlash va saqlash, yaratish, o'zgartirish va sinovdan o'tkazishni ta'minlovchi boshqaruv tizimini o'z ichiga oladi. DBMS ichki dasturlash tillari to'rtinchi avlod tillari (C, C ++, Paskal, Object Pascal). Ma'lumotlar bazasi tillari yordamida dasturlar, ma'lumotlar bazalari va foydalanuvchi interfeysi, shu jumladan ekran shakllari, menyular, hisobotlar yaratiladi. Agar tahlilchi, ma'lum bir MBBT bilan ishlash zarurati tug'ilsa, xususan, ma'lumotni ishlab chiqarish vositasi muhitiga ma'lumotlarni eksport qilishda ushbu MBBT xususiyatlarini o'rganishi kerak. Masalan, FoxPro ma'lumotlar bazasida barcha jadvallar va ma'lumotlar ko'rinishlari jismonan bitta loyihada birlashtirilgan alohida fayllarda saqlanadi. Access-da barcha ma'lumotlar bazalari jadvallari bitta faylda saqlanadi. Muayyan ma'lumotlar bazasi bilan ishlash uchun, shu jumladan tahlil qilish uchun tahlilchi barcha jadvallar va

ularning tuzilishlarini (atributlar, ma'lumotlar turlari) tavsifini, jadvaldagi yozuvlar sonini, shuningdek jadvallar o'rtasidagi munosabatlarni bilishi maqsadga muvofiqdir. Ba'zan bu maqsadda ma'lumotlar lug'ati ishlatiladi. Ma'lumotlar bazalariga, shuningdek, ma'lumotlar bazasiga quyidagi talablar qo'yiladi:

1. yuqori ishlash;
2. ma'lumotlarni yangilash qulayligi;
3. ma'lumotlar mustaqilligi;
4. ko'p foydalanuvchi ma'lumotlaridan foydalanish imkoniyati;
5. ma'lumotlar xavfsizligi;
6. ma'lumotlar bazasini qurish va ishlashini standartlashtirish (amalda ma'lumotlar bazasi);
7. tegishli fan sohasidagi ma'lumotlarni namoyish qilishning etarliligi;
8. do'stona interfeys.

Tez javob vaqtlari tezkor javob vaqtlarini anglatadi. Ma'lumotlar bazasi so'ralgan vaqtdan ma'lumot qabul qilingan paytgacha bo'lgan qisqa vaqt.

Ma'lumotlar mustaqilligi - bu foydalanuvchilarning qarashlarini o'zgartirmasdan ma'lumotlar bazasining mantiqiy va jismoniy tuzilishini o'zgartirish qobiliyati.

Ma'lumotlar mustaqilligi ma'lumotlar bazasi tarkibidagi minimal o'zgarishlarni, ma'lumotlarga kirish strategiyasini va manbaning o'zi tuzilishini o'zgartirilishini ta'minlaydi. Ushbu o'zgartirishlar ma'lumotlar bazasini kontseptual va mantiqiy loyihalash bosqichlarida, fizik dizayn bosqichida minimal o'zgarishlarni ta'minlagan holda ko'zda tutilishi kerak.

Ma'lumot xavfsizligi bu ma'lumotlarni qasddan yoki bila turib sirni buzish, buzish yoki yo'q qilishdan himoya qilishdir. Xavfsizlik ikkita tarkibiy qismdan iborat: yaxlitlik va ruxsatsiz kirishdan ma'lumotlarni himoya qilish.

Ma'lumotlar yaxlitligi - saqlanadigan ma'lumotlarning texnik nosozliklar, tizim xatolari va foydalanuvchilarning xatolari bilan bog'liq bo'lgan yo'q qilinish va yo'q qilinishga qarshilik.

Ma'lumotlar yaxlitligi - ma'lumotlarning aniqligi va ishonchliligi. Ma'lumotlar yaxlitligi quyidagilarni o'z ichiga oladi: noto'g'ri kiritilgan ma'lumotlar yo'qligi,

ma'lumotlar bazasini yangilashda xatolardan himoya qilish; turli jadvallardan tegishli ma'lumotlarni o'chirish (yoki kaskadli o'chirish) mumkin emasligi; texnik nosozliklar yuz berganda ma'lumotlar xavfsizligi (ma'lumotlarni tiklash qobiliyati) va boshqalar.

Ma'lumotni ruxsatsiz kirishdan himoya qilish ma'lum bir ma'lumotlar bazasiga kirishni cheklashni o'z ichiga oladi va xavfsizlik choralarini joriy etish yo'li bilan ta'minlanadi: turli foydalanuvchilarning funktsiyalariga va xizmat vazifalariga qarab ma'lumotlardan foydalanish huquqlarini farqlash; parol bilan himoyani joriy etish; ko'rinishlar yordamida, ya'ni. Asl nusxadan olingan va ma'lum foydalanuvchilarga muayyan muammolarni hal qilish uchun mo'ljallangan jadvallar.

Standartlashtirish ma'lum bir MBBT avlodlarining uzluksizligini ta'minlaydi, bir xil avlod ma'lumotlar bazalari ma'lumotlar bazalarining bir xil va turli xil ma'lumot modellari bilan o'zaro ta'sirini soddalashtiradi.

Ma'lumotlar bazasi ma'lumotlar bazasiga so'rovlarni ko'rib chiqish va javob olish uchun javobgardir. Ma'lumotni saqlash usullari boshqacha bo'lishi mumkin: ma'lumotlar modeli ham relatsion, ham ko'p o'lchovli, tarmoq yoki ierarxik bo'lishi mumkin.

2. Ma`lumotlar turlari

Ma'lumotlar turlarini Qanday ma'lumotlar bo'lishi mumkin? Bir nechta tasniflash quyida keltirilgan.

Nisbiy ma'lumotlar - bu relyatsion ma'lumotlar bazalari (jadvallar) dan olingan ma'lumotlar. Ko'p o'lchovli ma'lumotlar bu OLAP kublarida berilgan ma'lumotlar. O'lchov yoki o'q - ko'p o'lchovli ma'lumotlarda - bu ko'p o'lchovli ma'lumotlar bazasini tuzishga imkon beradigan bir xil turdagi ma'lumotlar to'plamidir. Muammoni yechishda ularning qiymatlari barqarorligi mezoniga ko'ra ma'lumotlar quyidagicha bo'lishi mumkin.

- o'zgaruvchilar;
- doimiy;
- shartli doimiy.

O'zgaruvchan ma'lumotlar - bu muammoni hal qilish jarayonida uning qiymatlarini o'zgartiradigan ma'lumotlar.

Doimiy ma'lumotlar bu muammoni yechishda o'z qadriyatlarini saqlaydigan ma'lumotlar (matematik konstantalar, statsionar jismlarning koordinatalari) va tashqi omillarga bog'liq emas.

Shartli doimiy ma'lumotlar - bu ba'zan uning qiymatlarini o'zgartirishi mumkin bo'lgan ma'lumotlar, ammo bu o'zgarishlar muammoni hal qilish jarayoniga bog'liq emas, lekin tashqi omillar bilan belgilanadi.

Ma'lumotlar, ular bajaradigan funktsiyalarga qarab, ma'lumot, operatsion, arxiv bo'lishi mumkin. Davr ma'lumotlari va nuqta ma'lumotlari o'rtasida farqni aniqlash kerak. Ushbu farqlar sotib olish tizimini loyihalashda, shuningdek o'lchash jarayonida muhimdir: davr uchun ma'lumotlar, nuqta ma'lumotlari.

Davr ma'lumotlari ma'lum vaqt oralig'ini tavsiflaydi. Bir davr uchun ma'lumotlarga misol bo'lishi mumkin: korxonaning oylikdagi foydasi, oylik o'rtacha harorat. Nuqta ma'lumotlari ma'lum bir vaqtning o'zida o'zgaruvchining qiymatini anglatadi. Nuqtali ma'lumotlarga misol: oyning birinchi kunidagi hisob qoldig'i, ertalab sakkizda harorat. Ma'lumotlar dastlabki va ikkinchi darajali. Ikkilamchi ma'lumotlar - bu dastlabki ma'lumotlarga nisbatan qo'llanilgan ma'lum hisob-kitoblar natijasida olingan ma'lumotlar. Ikkilamchi ma'lumotlar, qoida tariqasida, saqlanadigan ma'lumot miqdorini ko'paytirish orqali foydalanuvchining so'roviga tezkor javob olishga olib keladi.

Keng ma'noda ma'lumotlar bu faktlar, matn, grafika, rasmlar, tovushlar, analog yoki raqamli video segmentlardir.

O'lchovlar, tajribalar, arifmetik va mantiqiy operatsiyalar natijasida ma'lumotlarni olish mumkin. Ma'lumotlar saqlash, uzatish va qayta ishlash uchun mos shaklda taqdim etilishi kerak. Boshqacha qilib aytganda, ma'lumotlar - bu ma'lumot etkazib beruvchilar tomonidan ta'minlanadigan va iste'molchilar tomonidan ma'lumotlardan ma'lumotlarni shakllantirish uchun foydalaniladigan xom ashyo.

Ma'lumotlar to'plami va ularning xususiyatlari

2.2.1-jadvalda ma'lumotlar bazasini ifodalovchi ikki o'lchovli jadval keltirilgan.

Atributlar obyektlar				
Mijoz IDsi	**Yoshi**	**Oilaviy ahvoli**	**Daromad**	**Sinf**
1	18	Single	125	1
2	22	Married	100	1
3	30	Single	70	1
4	32	Married	120	1
5	24	Divorced	95	2
6	25	Married	60	1
7	32	Divorced	220	1
8	19	Single	85	2
9	22	Married	75	1
10	40	Single	90	2

2.2.1-jadval. "Ob'yekt-atribut" ikki o'lchovli jadval

Jadvalning gorizontal tomonida ob’yektning atributlari yoki uning belgilari joylashgan. Vertikal jadvallar ob’yektlardir. Ob'yekt atributlar to'plami sifatida tavsiflanadi. Ob'yekt yozuv, voqea, misol, jadval qatori va boshqalar sifatida ham tanilgan. Atribut - bu ob’yektni tavsiflovchi xususiyat.

Masalan: insonning ko'z rangi, suv harorati va boshqalar. Atribut shuningdek o'zgaruvchan, jadval maydoni, o'lchov, xarakterli deb ham ataladi. Kontseptsiyalarni tezlashtirish natijasida, ya'ni umumiy kategoriyalardan ma'lum qiymatlarga o'tish, o'rganilayotgan kontseptsiyaning o'zgaruvchilar to'plami olinadi.

O'zgaruvchi - bu barcha o'rganilayotgan ob’yektlar uchun umumiy bo'lgan, namoyon bo'lishi ob’yektdan ob’yektga o'zgarishi mumkin bo'lgan xususiyat yoki xususiyatdir. O'zgaruvchining qiymati xususiyatning namoyon bo'lishi. Ma'lumotni tahlil qilganda, qoida tariqasida, bizni qiziqtirgan barcha ob’yektlar to'plamini ko'rib chiqishning imkoni yo'q. Juda katta hajmdagi ma'lumotlarni o'rganish qimmatga tushadi va ko'p vaqt talab etadi va muqarrar ravishda inson xatolariga olib keladi.

Butun aholining ma'lum bir qismini, ya'ni namunani ko'rib chiqish va unga asoslanib biz uchun qiziq ma'lumotni olish kifoya. Shu bilan birga, tanlov hajmi

umumiy populyatsiyada aks ettirilgan ob’yektlarning xilma-xilligiga bog'liq bo'lishi kerak. Namuna turli xil kombinatsiyalar va populyatsiya elementlarini aks ettirishi kerak.

Umumiy to'plam (populyatsiya) - tadqiqotchini qiziqtirgan barcha o'rganilayotgan ob’yektlar to'plami.

Namuna (Sample) - ma'lum populyatsiyaning xususiyatlari va xususiyatlari to'g'risida tadqiq qilish va xulosalar chiqarish uchun ma'lum usulda tanlangan qismi.

Parametrlar - umumiy populyatsiyaning raqamli xususiyatlari.

Statistikalar - namunaning raqamli xarakteristikalari.

Tadqiqotlar ko'pincha farazlarga asoslangan. Gipotezalar ma'lumotlar bilan tasdiqlangan.

Gipoteza - bu ob’yekt tomonidan tekshirilishi kerak bo'lgan ob’yektlar to'plamining parametrlariga oid taxmin.

Gipoteza - bu turli xil empirik faktlar orasidagi bog'lanish uchun yoki bir faktni yoki bir guruh faktlarni tushuntirish uchun xizmat qiladigan qisman asoslangan bilimlar shakli.

Gipotezaga misol: umr ko'rish davomiyligi va ovqatlanish sifati o'rtasida bog'liqlik mavjud. Bunday holda, tadqiqotning maqsadi ma'lum bir o'zgaruvchining o'zgarishini, bu holda umr ko'rish davomiyligini tushuntirish bo'lishi mumkin. Aytaylik, bog'liq bo'lgan o'zgaruvchi (umr ko'rish davomiyligi) mustaqil o'zgaruvchilar bo'lgan ba'zi sabablarga (oziq-ovqat sifati, turmush tarzi, yashash joyi va boshqalar) qarab o'zgaradi. Ammo, o'zgaruvchi dastlab bog'liq yoki mustaqil emas. Muayyan gipoteza aniqlangandan keyin shunday bo'ladi. Bir gipotezada bog'liq bo'lgan o'zgaruvchi boshqasida mustaqil bo'lishi mumkin.

O'lchash - bu ma'lum bir qoidaga muvofiq o'rganilayotgan ob'yektlarning xususiyatlariga raqamlar berish jarayoni.

Ma'lumotni tayyorlash jarayonida ob'yektning o'zi emas, balki uning xususiyatlari o'lchanadi.

Miqyos bu ob’yektlarga raqamlar berish qoidasi. Ma'lumotni qazib olishning ko'plab vositalari, boshqa manbalardan ma'lumotlarni import qilishda, har bir

o'zgaruvchiga masshtab turini tanlashni yoki kirish va chiqish o'zgaruvchisi (ramziy, raqamli, diskret va uzluksiz) uchun ma'lumot turini tanlashni taklif qiladi. Bunday vositadan foydalanuvchi ushbu tushunchalar bilan tanishishi kerak. O'zgaruvchilar raqamli yoki ramziy bo'lishi mumkin. O'z navbatida sonli ma'lumotlar diskret va doimiy bo'lishi mumkin.

Diskret ma'lumotlar bu xususiyatlar qiymatidir, ularning umumiy soni cheksiz yoki cheksizdir, lekin ularni sonlardan cheksizgacha bo'lgan tabiiy sonlar yordamida hisoblash mumkin. Diskret ma'lumotlarga misol. Trolleybus yo'nalishining davomiyligi (kurs davomiyligi variantlari soni): 10, 15, 25 daqiqa.

Uzluksiz ma'lumotlar - ularning qiymatlari ma'lum vaqt oralig'ida istalgan qiymatni olishi mumkin bo'lgan ma'lumotlar. Uzluksiz ma'lumotlarni o'lchash katta aniqlikni talab qiladi. Uzluksiz ma'lumotlarga misol: harorat, balandlik, og'irlik, uzunlik va boshqalar.

Shkalalar - O'lchov birliklarining besh turi mavjud: nominal, tartibli, oraliq, nisbiy va dikotomik.

Nominal shkala - faqat toifalarni o'z ichiga olgan shkala; undagi ma'lumotlarga buyurtma berish mumkin emas, ular bilan arifmetik amallarni bajarish mumkin emas.

Nominal shkala nomlar, toifalar, ob’yektlarni tasniflash va saralash uchun nomlardan yoki ba'zi bir mezonga ko'ra kuzatuvlardan iborat.

Bunday ko'lamga misol: kasblar, yashash joyi, oilaviy ahvol.

Ushbu masshtab uchun faqat shunday operatsiyalar qo'llaniladi: teng (=), teng emas (\=).

Tartibli shkala bu ob’yektlarga nisbiy pozitsiyasini ko'rsatadigan raqamlar berilgan, ammo ular orasidagi farqning kattaligini anglatmaydigan o'lchov.

O'lchov shkalasi o'zgaruvchilarning qiymatlarini tartiblashtirishga imkon beradi. Tarkibiy o'lchovdagi o'lchovlar faqat kattalik tartibiga oid ma'lumotlarni o'z ichiga oladi, ammo "bitta qiymat boshqasidan kattaroq" yoki "u boshqasidan qanchalik kam" deb aytishga imkon bermaydi. Bunday shkalaga misol: musobaqada

jamoa olgan joy (1, 2, 3), talabalar ko'rsatkichi reytingida (1, 23, va hokazo), lekin bitta talabaning qanchalik muvaffaqiyatli ekanligi noma'lum. ikkinchisi, faqat uning tartib raqami ma'lum. Ushbu masshtab uchun faqat shunday operatsiyalar qo'llaniladi: teng (=), ((\=) ga teng bo'lmagan, (>) dan katta, (<) dan kichik.

Interval shkalasi - bu qiymatlar o'rtasidagi farqlarni hisoblash mumkin bo'lgan shkala, ammo ularning o'zaro aloqalari ma'nosizdir. Ushbu shkala ikki qiymat o'rtasidagi farqni topishga imkon beradi, nominal va tartibli o'lchovlarning xususiyatlariga ega, shuningdek, atributdagi miqdoriy o'zgarishlarni aniqlashga imkon beradi. Bunday o'lchovning misoli: ertalab dengiz suvining harorati - 19 daraja, kechqurun - 24, ya'ni. kechqurun 5 darajaga ko'tariladi, lekin uni 1,26 baravar yuqori deyish mumkin emas. Nominal va tartib o'lchovlari diskretdir va interval shkalasi doimiy bo'lib, bu xususiyatni aniq o'lchash va qo'shish, ayirish, ko'paytirish, bo'lish kabi arifmetik amallarni bajarishga imkon beradi. Ushbu masshtab uchun faqat shunday operatsiyalar qo'llaniladi: teng (=), teng bo'lmagan (\ =), kattaroq (>), kichik (<), qo'shish (+) va ayirish (-) operatsiyalari. Nisbiy shkala (nisbat shkalasi) - bu ma'lum bir yo'nalishda bo'lgan va shkala qiymatlari o'rtasidagi munosabatlar mumkin bo'lgan shkala. Bunday o'lchovga misol: yangi tug'ilgan chaqaloqning vazni (4 kg va 3 kg). Birinchisi 1,33 marta og'irroq. Supermarketda kartoshkaning narxi bozordagi narxdan 1,2 baravar yuqori. Nisbiy va oraliq o'lchovlar raqamli hisoblanadi. Ushbu masshtab uchun faqat shunday amallar qo'llaniladi: teng (=), teng bo'lmagan (\ =), (>) dan katta, (<) dan kichik, qo'shish (+) va ayirish (-), ayirish (*) va bo'linish (/) ...

Dixotomoz shkala - faqat ikki toifani o'z ichiga olgan shkala.

Bunday shkalaga misol - jins (erkak va ayol). Turli xil ob'yektlarning xususiyatlarini o'lchash uchun turli xil o'lchovlardan foydalanishga misol 2.2.2-jadvalda keltirilgan ma'lumotlar jadvalida keltirilgan.

Ob`yekt raqami	Kasbi(nominal shkalasi)	O'rtacha ball (oraliq shkala)	Ta'lim (tartib bo'yicha)
1	Chilangar	22	O`rta
2	Olim	55	Oliy
3	O`qituvchi	47	Oliy

2.2.2-jadval. Turli xil ob'yektlarning xususiyatlarini ko'plab o'lchash

Bitta tizimning xususiyatlarini o'lchash uchun turli xil o'lchovlardan foydalanishga misol, bu holda harorat sharoitlari 2.2.3-jadvalda keltirilgan ma'lumotlar jadvalida keltirilgan.

O'lchov sanasi	Bulutlilik (nominal shkalasi)	Ertalab 8 da harorat (oraliq shkalada)	Shamol kuchi (tartibga solish shkalasi)
1 sentabr	Bulutli	22 C	Kuchli shamol
2 sentabr	Asosan bulutli	17 C	Shamol kuchsiz
3 sentabr	Ochiq	23 C	Shamol juda kuchli

2.2.3-jadval. Bitta tizimning xususiyatlarini bir necha marta o'lchash

Xulosa. Ma'ruzaning ushbu qismida ma'lumotlar, ob’yekt va atribut tushunchalari va ularning xususiyatlarini ko'rib chiqdik.

Shuningdek, biz tarozi turlarini muhokama qildik. Nominal shkala ob’yektlarni yoki kuzatuvlarni sifat xususiyatlari jihatidan tavsiflaydi. Bir qadam oldinga, ma'lum bir xususiyatga ko'ra kuzatuvlar yoki ob’yektlarni tartibga solish imkonini beradigan tartibli tarozilar mavjud. Interval va nisbiy o'lchovlar yanada murakkabroq bo'lib, unda xususiyatning miqdoriy qiymatini aniqlash mumkin.

3. Ma`lumotlarni saqlash formatlari

Yozuvlardan iborat ma'lumotlar. Eng keng tarqalgan ma'lumotlar bu yozuvlardan iborat ma'lumotlar (rekord ma'lumotlar). Bunday ma'lumotlar to'plamiga misollar jadval jadvallari, matritsa ma'lumotlari, hujjatli ma'lumotlar, tranzaktsion yoki operatsion ma'lumotlardir.

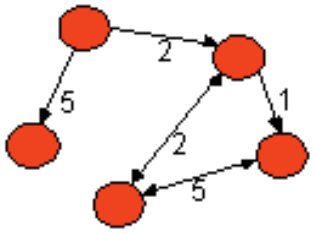
jadval ma'lumotlari - yozuvlardan iborat ma'lumotlar, ularning har biri belgilangan atributlar to'plamidan iborat.

Tranzaktsion ma'lumotlar - bu ma'lumotlarning maxsus turi bo'lib, unda har bir yozuv bir qator qiymatlarni o'z ichiga oladi. Do'kon mijozlari tomonidan xaridlar ro'yxatini o'z ichiga olgan tranzaksiya ma'lumotlar bazasining namunasi 2.3.1.-jadvalda.

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

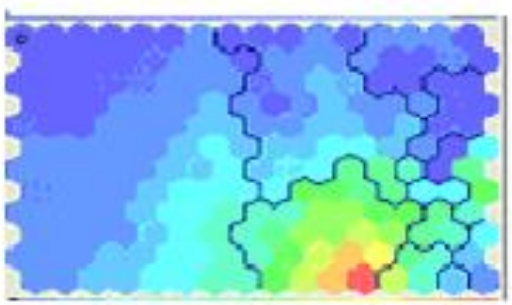
2.3.1.-jadval Tranzaktsion ma'lumotlarga misol

Grafik ma'lumotlar. Grafik ma'lumotlarga misollar: WWW ma'lumotlari; molekulyar tuzilmalar; grafikalar 2.3.1-shakl); kartalar.



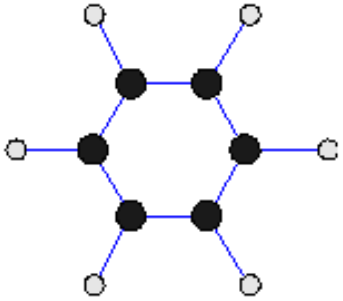
Shakl: 2.3.1. Kohonen xaritasi ma'lumotlari namunasi

Masalan, xaritalardan foydalanib, ob’yektlardagi o'zgarishlarni vaqt va makonda kuzatish, ularning samolyotda yoki kosmosda tarqalish xususiyatini aniqlash mumkin. Ma'lumotlarning grafik taqdimotining afzalligi, masalan, jadval jadvallariga qaraganda, uni idrok qilishning osonligi. Kohonen xaritasi bo'lgan xaritaning misoli (bizning kursimizning ma'ruzalaridan birida muhokama qilinadigan neyron tarmoqlarining modeli) shakl. 2.3.2.



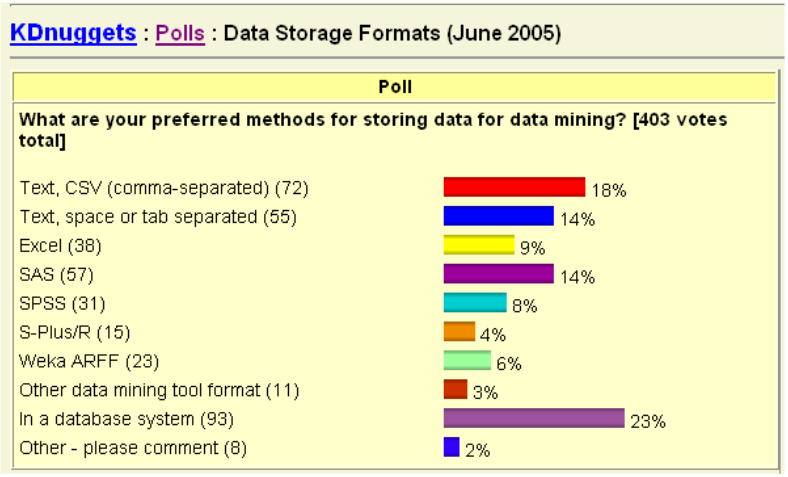
Shakl: 2.3.2. Kohonen xaritasi ma'lumotlari namunasi

Kimyoviy ma'lumotlar. Kimyoviy ma'lumotlar bu ma'lumotlarning maxsus turi. Bunday ma'lumotlarga misol: Benzol molekulasi: C6H6 (2.3.3-shakl)



Shakl: 2.3.3. Kimyoviy ma'lumotlarga misol

Kdnuggets veb-saytidagi www.kdnuggets.com (2004 yil aprel) "Tahlil qilinadigan ma'lumotlarning turlari" so'roviga ko'ra, respondentlarning eng ko'p qismi "tekis" va o'zaro bog'liq jadvallardagi ma'lumotlarni (mos ravishda 26% va 24%) tahlil qiladi. Vaqt seriyalari (14%) va matn shaklida ma'lumotlar mavjud (11%). Qolgan tahlil qilingan ma'lumotlar turlari kamayish tartibida: veb-tarkib, XML, grafika, audio, video va boshqalar. Bu yerda va quyidagi ma'ruzalarda siz ma'lumot ishlab chiqarish sanoatida eng nufuzli va taniqli saytlardan biri sifatida tan olingan Kdnuggets saytida o'tkazilgan so'rovlarning natijalarini topasiz. Zamonaviy dunyoda ma'lumotlarning asosiy xususiyatlaridan biri shundaki, ular juda ko'p. Ma'lumotlar bilan ishlashning to'rt jihati mavjud: ma'lumotlarni aniqlash, hisoblash, boshqarish va ishlov berish (to'plash, uzatish va boshqalar). Ma'lumotni boshqarish paytida "fayl" turidagi ma'lumotlarning tuzilishi qo'llaniladi. Fayllar har xil formatda bo'lishi mumkin. Yuqorida ta'kidlab o'tilganidek, Data Mining vositalarining aksariyati turli manbalardan ma'lumotlarni import qilish, shuningdek olingan ma'lumotlarni turli formatlarda eksport qilish imkonini beradi. Tajribalar uchun ma'lumotlarni yagona formatda saqlash qulay. Ba'zi bir ma'lumot ishlab chiqarish vositalarida ushbu protseduralar ma'lumotlarni import / eksport deb nomlanadi, boshqalari esa har xil ma'lumot manbalarini to'g'ridan-to'g'ri ochish va Data Mining natijalarini taklif qilingan formatlardan biriga saqlash imkonini beradi. "Ma'lumotni saqlash formatlari" so'roviga ko'ra eng keng tarqalgan formatlar shakl – 2.3.4.



Shakl: 2.3.4. Ma'lumotni saqlashning eng keng tarqalgan formatlari

Respondentlarning eng ko'p soni (23%) ma'lumotni ular ishlatadigan ma'lumotlar bazasi formatida saqlashni afzal ko'rishadi. Matnda, CSV formatida - 18%, respondentlarning 14% har bir ma'lumotni Matn, bo'sh joy yoki yorliqda ajratilgan va SAS formatida saqlaydi; Excel formatida - 9%, SPSS - 8%, S-Plus / R - 4%, Weka ARFF - 6%, Data Mining vositalarining boshqa formatlarida - 2%. So'rov natijalaridan ko'rinib turibdiki, Ma'lumotlar bazalari Data Mining uchun eng keng tarqalgan ma'lumotlarni saqlash formatidir.

4. Metama`lumotlar

Metadata(Metama`lumotlar) bu ma'lumotlar haqidagi ma'lumotlar. Metadata o'z ichiga olishi mumkin: kataloglar, ma'lumotnomalar, registrlar. Metadata ma'lumotlarning tarkibi, tarkibi, holati, kelib chiqishi, joylashuvi, sifati, taqdimot shakllari va shakllari, kirish, olish va ulardan foydalanish shartlari, mualliflik huquqi, ma'lumotlarga nisbatan mulk huquqi va boshqalar to'g'risidagi ma'lumotlarni o'z ichiga oladi.

Metadata - bu ma'lumotlar omborini boshqarishda muhim tushunchadir. Rezervuarni boshqarish uchun ishlatiladigan metadata uni sozlash va undan foydalanish uchun zarur bo'lgan ma'lumotlarni o'z ichiga oladi. Ish meta-ma’lumotlari va operatsion meta-ma’lumotlarni farqlash. Ish metama'lumotlarida biznes atamalari va ta'riflari, ma'lumotlarga egalik va saqlash haqlari mavjud.

Operatsion metadata - bu ma'lumotlar ombori ishlashi davomida to'plangan ma'lumotlar:

- uzatilgan va o'zgartirilgan ma'lumotlarning kelib chiqishi;
- ma'lumotlardan foydalanish holati (faol, arxivlangan yoki o'chirilgan);
- foydalanish statistikasi, xato xabarlari va boshqalar kabi monitoring ma'lumotlari

Repository metadata odatda omborxonada joylashgan. Bu metadata ma'lumotlarini omborni loyihalash, o'rnatish, ishlatish va boshqarishdagi turli xil vositalar va jarayonlar o'rtasida almashish imkonini beradi.

Xulosa. Ma'ruzada ma'lumotlar, ob'yektlar va atributlar tushunchasi, ularning xususiyatlari, o'lchov turlari, ma'lumotlar bazasi tushunchasi va uning turlari ko'rib chiqildi. Mumkin bo'lgan ma'lumotlarni saqlash formatlari tavsiflangan. Ma'lumotlar bazasi tushunchalari, ma'lumotlar bazasini boshqarish tizimi, metadata.

Nazorat savollari:

1. Ma`lumotlar bazasi yaratilish tarixi haqida nimani bilasiz?
2. Ma`lumot turlariga misollar keltiring.
3. MBBT nima va uning qanday turlari mavjud?
4. Ma`lumot formatlari deganda nimani tushunasiz?

3-MAVZU

DATA MINING METODLARI VA BOSQICHLARI

Reja:

1.Data Mining metodlari

2.Data Mining bosqichlarini tasniflash

**Mashg`ulot maqsadi:** *Mashg`ulotda Data Mining bosqichlari va ushbu bosqichlarda amalga oshirilgan harakatlar tasvirlangan. Data mining usullarining taniqli tasniflari ko'rib chiqildi. Xususiyatlariga asoslangan ba'zi usullarning qiyosiy xarakteristikasi berilgan.*

Tayanch iboralar: *Data Mining, usul, tahlil, algoritm, sun'iy neyron tarmoqlar, qo'llab-quvvatlash vektori, evolyutsion dasturlash, usul, yo'l, algoritm, naqsh, erkin izlash, naqsh, taxminiy modellashtirish, istisnolarni tahlil qilish, qonun, yaqinlik, trend, o'zgaruvchanlik, bashorat model, natija, prognoz, tendentsiya, og'ish, ma'lumotlarni tozalash, ma'lumotlarni saqlash, shablonni distillash, terminal tepasi, o'lchovlilik, bilimlarni ekstraksiya qilish, omillarni tahlil qilish, dinamik model, havolalarni tahlil qilish, dispersiyani tahlil qilish, assotsiativ xotira, o'z-o'zini tashkil etuvchi xarita.*

1.Data Mining metodlari

Data Mining-ning asosiy xususiyati bu matematik vositalarning keng to'plamidir (klassik statistik tahlildan yangi kibernetik usullargacha) va axborot texnologiyalarining eng so'nggi yutuqlari. Data Mining texnologiyasi qat'iy rasmiylashtirilgan usullar va norasmiy tahlil usullarini uyg'un ravishda birlashtiradi, ya'ni. miqdoriy va sifatli ma'lumotlarni tahlil qilish.

Ma'lumotlarni qidirish usullari va algoritmlari quyidagilarni o'z ichiga oladi: sun'iy neyron tarmoqlari, qaror daraxtlari, ramziy qoidalar, eng yaqin qo'shni va k-yaqin qo'shni usullari, qo'llab-quvvatlovchi vektor mashinalari, Bayesiya tarmoqlari, chiziqli regressiya, korrelyatsiya-regressiya tahlili; klasterli tahlilning ierarxik usullari, klasterli tahlilning ierarxik bo'lmagan usullari, shu jumladan k-

vositalari va k-median algoritmlari; birlashma qoidalarini, shu jumladan Apriori algoritmini topish usullari; chegaralangan ro'yxatga olish usuli, evolyutsion dasturlash va genetik algoritmlar, ma'lumotlarni vizuallashtirishning turli usullari va boshqa ko'plab usullar.

Data Mining texnologiyasida qo'llaniladigan tahliliy usullarning aksariyati taniqli matematik algoritmlar va usullardir. Ularning qo'llanilishida yangilik - bu dasturiy ta'minot va dasturiy ta'minotlarning paydo bo'lishi imkoniyatlari tufayli ma'lum bir muammolarni hal qilishda ulardan foydalanish imkoniyati. Shuni ta'kidlash kerakki, Data Mining usullarining aksariyati sun'iy aql nazariyasi doirasida ishlab chiqilgan.

Usul (method) – bu norma yoki qoida, ma'lum bir usul, usul, nazariy, amaliy, kognitiv, boshqaruv xarakteridagi muammolarga echimlarni qabul qilish.

Algoritm tushunchasi elektron kompyuterlar yaratilishidan ancha oldin paydo bo'lgan. Hozirgi vaqtda algoritmlar inson faoliyatining turli sohalarida amaliy va nazariy muammolarni hal qilish uchun asos bo'lib, ularning aksariyati echimi kompyuter yordamida ta'minlangan vazifalardir.

Algoritm - bu dastlabki ma'lumotlarni kerakli natijaga aylantiradigan harakatlar ketma-ketligi (bosqichlari) uchun aniq natija.

2. Data Mining bosqichlarini tasniflash

Data Mining ikki yoki uch bosqichdan iborat bo'lishi mumkin:

1-Bosqich. Qonuniyatlarni identifikatsiyalash(oddiy qidiruv).

2-Bosqich. Noma'lum qiymatlarni bashorat qilish uchun (aniqlangan modellash) aniqlangan qiymatlardan foydalanish.

3-Bosqich. Istisno tahlil – oldindan aniqlangan anomaliyalarni aniqlash va tushuntirish uchun mo'ljallangan.

Ushbu bosqichlarga qo'shimcha ravishda, ba'zida bepul qidirish bosqichidan keyin tekshirish bosqichi joriy etiladi. Tasdiqlashning maqsadi topilgan naqshlarning ishonchliligini tekshirish. Biroq, biz birinchi bosqichning bir qismi sifatida tekshirishni ko'rib chiqamiz, chunki ko'plab usullarni, xususan, neyron

tarmoqlari va qaror daraxtlarini amalga oshirish, ma'lumotlarning umumiy to'plamini o'qitish va tekshirishga bo'lishni ta'minlaydi va ikkinchisi olingan natijalarning ishonchliligini tekshirishga imkon beradi. Shunday qilib, Data Mining jarayoni quyidagi ketma-ket bosqichlar bilan ifodalanishi mumkin:

Oddiy qidiruv (shu jumladan validatsiya), Oldindan modellashtirish, Istisnolar tahlili

1. Oddiy qidirish (Discovery)

Oddiy qidirish bosqichida yashirin naqshlarni topish uchun ma'lumotlar to'plamini o'rganish amalga oshiriladi. Naqsh turiga oid dastlabki farazlar bu erda aniqlanmagan.

Qonuniyat - bu sodir bo'lish, turli hodisalar yoki jarayonlarning rivojlanish bosqichlari va shakllarini belgilovchi zaruriy va doimiy takrorlanadigan munosabatlardir. Ushbu bosqichda Data Mining tizimi shablonlarni aniqlaydi, ular uchun OLAP tizimlarida, masalan, tahlilchi o'ylab topishi va ko'plab so'rovlarni yaratishi kerak. Bu erda tahlilchi bunday ishlardan ozod qilinadi - tizim unga shablonlarni qidirmoqda. Ushbu yondashuv, ayniqsa, juda katta ma'lumotlar bazalarida foydalidir, bu erda so'rovlarni yaratish orqali naqshni topish qiyin va bu ko'plab turli xil variantlarni sinab ko'rishni talab qiladi.

Oddiy qidirish quyidagi harakatlar bilan ta'minlanadi:

- shartli mantiq qonuniyatlarni aniqlash;
- assotsiativ mantiqning qonuniyatlarni aniqlash (assotsiatsiyalar va yaqinliklar);
- tendentsiyalar va o'zgarishlarni aniqlash.

Aytaylik, sizda mutaxassislik, ish staji, yoshi va istalgan ish haqi darajasi to'g'risidagi ma'lumotlarga ega bo'lgan yollash agentligining ma'lumotlar bazasi mavjud. Mustaqil ravishda so'rovlar topshirilgan taqdirda, tahlilchi quyidagi natijalarga erishishi mumkin: 25 yoshdan 35 yoshgacha bo'lgan mutaxassislar uchun o'rtacha ish haqi darajasi 1200 shartli birlikka teng. Bepul qidirish holatida tizim o'zi naqshlarni qidiradi, faqat maqsad o'zgaruvchini belgilash kerak. Naqshlarni qidirish natijasida tizim "agar ... keyin ..." mantiqiy qoidalar to'plamini hosil qiladi.

Masalan, quyidagi naqshlarni topish mumkin: "Agar yoshi <20 yil bo'lsa va kerakli ish haqi darajasi> 700 shart birligi bo'lsa, u holda 75% hollarda ariza beruvchi dasturchi sifatida ish qidirmoqda" yoki "Agar 35 yoshga to'lgan bo'lsangiz va xohlagan ish haqingiz darajasi> 1200 shart birligi bo'lsa, unda holatlarning 90% da murojaat etuvchi rahbarlik ishini qidirmoqda ». Ta'riflangan qoidalardagi maqsad o'zgaruvchisi - bu kasb.

Boshqa maqsadli o'zgaruvchini belgilashda, masalan, yoshi, biz quyidagi qoidalarni olamiz: "Agar ariza beruvchi boshqaruv ishini qidirayotgan bo'lsa va uning tajribasi> 15 yil bo'lsa, unda 65% hollarda talabnoma beruvchining yoshi 35 yoshni tashkil etadi".

Oddiy qidirish bosqichida tavsiflangan harakatlar quyidagilar yordamida amalga oshiriladi:

- shartli mantiq qoidalarini tanishtirish (tasniflash va klasterlash muammolari, ob’yektlarning yaqin yoki o'xshash guruhlarini ixcham shaklda tavsiflash);
- assotsiativ mantiq qoidalarini tanishtirish (assotsiatsiya va ketma-ketlikdagi muammolar va ularning yordami bilan olingan ma'lumotlar);
- tendentsiyalar va tebranishlarni aniqlash (prognozlash muammosining dastlabki bosqichi).

Oddiy qidirish bosqichida qonuniyatlar ham tekshirilishi kerak, ya'ni. naqshlarni shakllantirishda ishtirok etmagan ma'lumotlarning qismlari bo'yicha ularning ishonchliligini tekshirish. Ma'lumotni o'quv va test to'plamlariga bo'lishning ushbu usuli ko'pincha neyron tarmoqlari va qaror daraxtlari usullarida qo'llaniladi va tegishli ma'ruzalarda tavsiflanadi.

2. Oldindan modellashtirish

Data Mining-ning ikkinchi bosqichi – bashoratni(oldindan) modellashtirish - birinchi bosqich natijalaridan foydalanadi. Bu erda topilgan naqshlar to'g'ridan-to'g'ri bashorat qilish uchun ishlatiladi.

Bashoratli modellashtirish quyidagi harakatlarni o'z ichiga oladi:

- noma'lum qiymatlarni bashorat qilish (natijani bashorat qilish);
- jarayonlarning rivojlanishini prognoz qilish (prognozlash).

Bashoratli modellashtirish jarayonida tasniflash va prognozlash muammolari hal qilinadi.

Tasniflash muammosini hal qilishda birinchi bosqich natijalari (qoida induktsiyasi) yangi ob’yektni ma'lum ishonch bilan ma'lum qadriyatlarga asoslangan, oldindan belgilangan sinflardan biriga tasniflash uchun ishlatiladi.

Prognozlash muammosini hal qilishda, maqsadli o'zgaruvchi (lar) ning noma'lum (etishmayotgan yoki kelajakda) qiymatlarini bashorat qilish uchun birinchi bosqich natijalari (trend yoki tebranishlarni aniqlash) ishlatiladi.

Birinchi bosqichning ko'rib chiqilgan misolini davom ettirib, quyidagi xulosaga kelishimiz mumkin.

Arizachining menejment bo'yicha ish qidirayotganini va uning tajribasi 15 yil ekanligini bilgan holda, 65% murojaat etuvchining 35 yoshga to'lganiga amin bo'lishi mumkin. Shu bilan bir qatorda, agar talabnoma beruvchining yoshi 35 yoshni tashkil etsa va kerakli ish haqi darajasi> 1200 shartni tashkil etsa, siz 90% murojaat etuvchini boshqaruv ishini qidirayotganiga amin bo'lishingiz mumkin.

Mantiq nuqtai nazaridan oddiy qidirish va bashoratli modellashtirishni taqqoslash

Oddiy qidirish umumiy naqshlarni ochib beradi. Bu tabiatda induktivdir. Ushbu bosqichda olingan naqshlar aniqdan umumiygacha shakllanadi. Natijada, ushbu sinfning alohida vakillarini o'rganish asosida ob’yektlarning ma'lum bir sinfi to'g'risida umumiy bilimga ega bo'lamiz.

Qoida: "Agar talabnoma beruvchining yoshi <20 yil bo'lsa va kerakli ish haqi darajasi> 700 shartni tashkil etsa, 75% hollarda murojaat etuvchi dasturchi sifatida ish qidirmoqda"

Xususan, ya'ni. "yoshi <20 yil" va "kerakli ish haqi darajasi> 700 shartli birliklar" sinfining ba'zi xususiyatlari haqida ma'lumot, biz umumiy haqida xulosa qilamiz, ya'ni: abituriyentlar - dasturchilar.

Bunga javoban, bashoratli modellashtirish deduktivdir. Ushbu bosqichda olingan naqshlar umumiydan o'ziga xos va yakka tartibda shakllanadi. Bu erda biz ob’yekt yoki ob’yektlar guruhi haqida yangi ma'lumotlarga ega bo'lamiz:

- tekshirilayotgan ob’yektlar qaysi sinfga tegishli ekanligini bilish;
- ob’yektlarning muayyan klassi doirasidagi amaldagi umumiy qoidani bilish.

Biz bilamizki, murojaat etuvchi boshqaruv ishini qidirmoqda va uning tajribasi> 15 yil, 65% murojaat etuvchining 35 yoshda ekanligiga amin bo'lishlari mumkin.

Ba'zi bir umumiy qoidalarga asoslanib, ya'ni: ariza beruvchining maqsadi - boshqaruv va uning tajribasi> 15 yil, biz bitta haqida xulosa qilamiz - arizachining yoshi 35 yosh.

Shuni ta'kidlash kerakki, olingan naqshlar, aniqrog'i, ularning dizayni shaffof bo'lishi mumkin, ya'ni. tahlilchi tomonidan sharhlanishi mumkin (yuqorida muhokama qilingan qoidalar) va "qora qutilar" deb nomlangan shaffof. Oxirgi dizaynning odatiy namunasi neyron tarmoqdir.

3. Istisnolarni tahlil qilish (sud-tibbiy tahlil)

Ma'lumotlar qidirishning uchinchi bosqichida topilgan naqshlarda aniqlangan istisnolar yoki anomaliyalar tahlil qilinadi. Ushbu bosqichda amalga oshirilgan harakatlar og'ishlarni aniqlashdir. Burilishlarni aniqlash uchun bepul qidirish bosqichida hisoblangan tezlikni aniqlash kerak. Yuqorida muhokama qilingan misollardan biriga qaytaylik.

"Agar yoshi 35 ga teng bo'lsa va talab qilinadigan ish haqi darajasi> 1200 shartli birlikka teng bo'lsa, 90% holatlarda murojaat etuvchi rahbarlik lavozimiga ish qidirmoqda" degan qoidani topdi. Savol tug'iladi - qolgan 10% ishlarni nima bilan bog'lash kerak?

Bu erda ikkita imkoniyat mavjud. Ulardan birinchisi, ba'zi bir mantiqiy tushuntirishlar mavjud bo'lib, ularni qoida sifatida shakllantirish mumkin. Qolgan 10% uchun ikkinchi variant - dastlabki ma'lumotlarda xatolar. Bunday holda, istisnolarni tahlil qilish bosqichida ma'lumotlarni tozalash sifatida foydalanish mumkin.

Nazorat savollari

1.Ma`lumotlarni qidirishning qanday usullarini bilasiz?

2.Algoritm nima?

3.Data Mining necha bosqichdan tashkil topishi mumkin va ular qaysilar?

4.Data Mining bosqichlariga misollar keltiring.

4-MAVZU

DATA MINING MASALALARI. AXBOROT VA BILIM.

Reja:

1.Ma'lumotlarni qidirish vazifalari

2.Axborot va bilim

**Mashg`ulot maqsadi:** *Mashg`ulotda Data Mining vazifalarining asosiy mohiyati va ularning tasnifi qisqacha bayon etilgan. "Axborot", "bilim", shuningdek ushbu tushunchalarni yaratish va taqqoslash tushunchalari batafsil ko'rib chiqiladi.*

Tayanch iboralar: *Ma'lumotlar, Data Mining muammosi, tasniflash, klasterlash, prognozlash, assotsiatsiya, vizualizatsiya, tahlil, taxmin qilish, ulanishni tahlil qilish, xulosa qilish, taqdimot, dasturiy ta'minot, ma'lumot, tasnif, ob'yekt, bayesiya tarmoqlari, indüksiyon, neyron tarmoq, bo'lim, o'z-o'zini xaritani, qidiruvni, algoritmni, ketma-ketlikni, ketma-ketlikni, ketma-ketlikni, assotsiatsiyani, naqshni, televizorni, mijozni, boshqaruvni, prognozni, og'ishni, aniqlashni, ma'lumotlarni tahlil qilishni, to'plamlarni, taxminiy, tahlilni, vizuallashtirishni, grafikani, 3 o'lchamli, xulosani, nazoratsiz o'rganishni tavsiflovchi , bashoratli vazifa, tavsiflovchi vazifa, bashoratli, bepul qidiruv, genetik algoritm, oqim, aloqa, qaror qabul qilish, axborot piramidasi, elektron biznes razvedkasi, bilim, nazorat, bashoratli modellashtirish, segmentatsiya, ma'lumotlar bazasi, kasb, firma, xarajatlar, talqin, xulosa, bilim, elektron hujjat aylanishi.*

4. Ma'lumotlarni qidirish vazifalari

Eslatib o'tamiz, Data Mining texnologiyasi andoza tushunchasiga asoslangan. Ko'zdan yashiringan ushbu anodzalarni aniqlash natijasida Data Mining vazifalari hal qilinadi. Ma'lumot qazib olishning ba'zi vazifalari odamlar uchun tushunarli bo'lgan shaklda ifodalanishi mumkin bo'lgan turli xil andozalarga mos keladi. Data Mining vazifalari ba'zida muntazamlik yoki texnikalar deb nomlanadi.

Data Mining-ga qanday vazifalarni kiritish kerakligi to'g'risida kelishuv yo'q. Vakolatli manbalarning aksariyati quyidagilarni o'z ichiga oladi: tasniflash,

klasterlash, prognozlash, birlashtirish, vizualizatsiya, og'ishlarni tahlil qilish va aniqlash, taxmin qilish, munosabatlarni tahlil qilish, xulosa qilish.

Keyingi ta'rifning maqsadi Data Mining-ning vazifalari haqida umumiy tushuncha berish, ularning ba'zilarini taqqoslash, shuningdek ushbu vazifalarni hal qilishning ba'zi usullarini taqdim etish. Ma'lumot qazib olishning eng keng tarqalgan vazifalari - tasniflash, klasterlash, birlashtirish, prognozlash va vizualizatsiya - keyingi ma'ruzalarda batafsil muhokama qilinadi. Shunday qilib, vazifalar ishlab chiqarilgan ma'lumotlarning turlariga qarab bo'linadi [18], bu Data Mining vazifalarining eng umumiy tasnifi. Data Mining muammolarini hal qilish usullari bilan batafsil tanishish kursning keyingi qismida taqdim etiladi.

Tasniflash(klassifikatsiya)

Qisqacha. Eng oddiy va keng tarqalgan Data Mining vazifasi. Tasniflash muammosini hal qilish natijasida o'rganilayotgan ma'lumotlar to'plamlari ob’yektlari guruhlarini tavsiflovchi belgilar topildi; ushbu asoslarda yangi ob’yekt u yoki bu sinfga tegishli bo'lishi mumkin.

Yechish usullari. Tasniflash muammosini hal qilish uchun quyidagi usullardan foydalanish mumkin: Yaqin qo'shnilar; k-yaqin qo'shni; Bayesian Tarmoqlari; qaror daraxtlarini kiritish; neyron tarmoqlari.

Klasterlash. Qisqacha. Klasterlash tasniflash g'oyasining mantiqiy davomidir. Bu vazifa yanada murakkab, klasterlashning o'ziga xos xususiyati shundaki, ob’yektlar sinflari oldindan belgilanmagan. Klasterlash ob’yektlarning guruhlarga bo'linishiga olib keladi.

Klaster muammosini hal qilish usuliga misol: neyron tarmoqlarining maxsus turini - o'z-o'zini tashkil etuvchi Kohonen xaritalarini nazoratsiz o'rganish.

Assotsiatsiya. Qisqacha. Assotsiatsiya qoidalarini topish muammosini hal qilish jarayonida ma'lumotlar bazasida bog'liq bo'lgan hodisalar o'rtasida shakllar mavjud. Assotsiatsiya va oldingi ikkita ma'lumot ishlab chiqarish vazifalari o'rtasidagi farq: shakllarni qidirish tahlil qilinadigan ob’yektning xususiyatlariga emas, balki bir vaqtning o'zida sodir bo'ladigan bir nechta hodisalar o'rtasida amalga

oshiriladi. Assotsiatsiya qoidalarini topish muammosini hal qilishning eng mashhur algoritmi bu Apriori algoritmi.

Navbat yoki ketma-ket assotsiatsiya.

Qisqacha. Muvofiqlik sizga bitimlar o'rtasidagi vaqtinchalik shakllarni topishga imkon beradi. Biror ketma-ketlikning vazifasi birlashma bilan o'xshashdir, ammo uning maqsadi bir vaqtning o'zida sodir bo'layotgan voqealar o'rtasida emas, balki vaqt bilan bog'liq bo'lgan hodisalar (ya'ni muayyan vaqt oralig'ida sodir bo'ladigan) o'rtasida qonuniyatlarni o'rnatishdir. Boshqacha qilib aytganda, ketma-ketlik vaqt bilan bog'liq voqealar zanjirining yuqori ehtimoli bilan belgilanadi. Aslida, assotsiatsiya nolga teng vaqt bosqichi bo'lgan ketma-ketlikning maxsus holatidir. Ushbu Data Mining vazifasi ketma-ket shakl vazifasi sifatida ham tanilgan. Tartibga solish qoidasi: X hodisadan keyin Y voqeasi ma'lum vaqtdan keyin sodir bo'ladi.

Misol. Kvartirani sotib olgandan keyin, 60% hollarda ijarachilar ikki hafta ichida muzlatgich sotib olishadi, va ikki oy ichida, 50% hollarda televizor sotib olinadi. Ushbu muammoning echimi marketing va menejmentda, masalan, mijozlar tsiklini boshqarishda keng qo'llaniladi (Customer Lifecycle Management).

Prognozlash. Tarixiy ma'lumotlarning xususiyatlariga asoslangan prognozlash muammosini hal qilish natijasida maqsadli raqamli ko'rsatkichlarning etishmayotgan yoki kelajakdagi qiymatlari baholanadi. Bunday muammolarni hal qilishda matematik statistika usullari, neyron tarmoqlar va boshqalar keng qo'llaniladi.

Og`ishlarni aniqlash, og'ish yoki tashqi tahlil. Ushbu muammoni hal qilishning maqsadi umumiy ma'lumot to'plamidan eng farq qiladigan ma'lumotlarni aniqlash va tahlil qilish, uncharacteristic shakl deb ataladigan narsalarni aniqlashdir.

Baholash. Hisoblash vazifasi xususiyatning doimiy qiymatlarini bashorat qilishgacha kamayadi.

Bog'lanishni tahlil qilish - ma'lumotlar bazasida bog'liqlikni topish vazifasidir.

Vizualizatsiya (Visualization, Graph Mining). Vizualizatsiya natijasida tahlil qilingan ma'lumotlarning grafik tasviri yaratiladi. Vizualizatsiya muammosini hal qilish uchun ma'lumotlarda shakl mavjudligini ko'rsatadigan grafik usullar qo'llaniladi. Vizualizatsiya texnikasiga misol sifatida 2D va 3D o'lchamlarda ma'lumotlarni taqdim etish mumkin.

Xulosa qilish - bu tahlil qilinadigan ma'lumotlar bazasidagi ob’yektlarning aniq guruhlarini tavsiflash vazifasi.

Data Mining vazifalarini klassifikatsiyalash. Strategiyalar bo'yicha tasnifga ko'ra, Data Mining vazifalari quyidagi guruhlarga bo'linadi:

- o'qituvchi bilan dars berish;
- o'qituvchisiz dars berish;
- va boshqalar.

Nazorat ostidagi o'quv toifasi quyidagi ma'lumotlarni ishlab chiqish vazifalari bilan ta'minlangan: tasniflash, baholash, prognozlash.

Nazorat qilinmagan o'quv toifasi klasterlash vazifasi bilan ifodalanadi.

Boshqa toifaga oldingi ikkita strategiyada ko'zda tutilmagan vazifalar kiradi.

Amaldagi modellarga qarab, Data Mining vazifalari tavsiflovchi va bashoratli bo'lishi mumkin. Ushbu turdagi modellar ma'lumotni ishlab chiqarish jarayoni haqidagi ma'ruzada batafsil tavsiflanadi. Ushbu tasnifga muvofiq, Data Mining vazifalari tavsiflovchi va bashoratli vazifalar guruhlari tomonidan taqdim etiladi.

Ta'rif vazifalarini hal qilish natijasida tahlilchi talqin qilish uchun mumkin bo'lgan ma'lumotlarni tavsiflovchi shablonlarni oladi. Ushbu vazifalar tahlil qilingan ma'lumotlarning umumiy tushunchasini tavsiflaydi, ma'lumotlarning informatsion, yakuniy, farqlovchi xususiyatlarini aniqlaydi. Ta'rif vazifalari tushunchasi ma'lumotlar to'plamlarini tavsiflash va taqqoslashni o'z ichiga oladi. Ma'lumotlar to'plamining xarakteristikasi ma'lumotlar to'plamining qisqa va qisqacha tavsifini beradi. Taqqoslash ikki yoki undan ortiq ma'lumotlar to'plamining qiyosiy tavsifini beradi

Prognoz ma'lumotlar tahlili - modellashtirishga, yangi yoki noma'lum ma'lumotlarning tendentsiyalari yoki xususiyatlarini bashorat qilishga asoslangan.

Ma'lumotlar konstruktsiyasining vazifalari quyidagilarga bo'linadi: tadqiqot va kashfiyot, bashorat qilish va tasniflash, tushuntirish va tavsiflash.

Avtomatik qidirish va toppish. Masalaning qo`yilishi: bozorning yangi segmentlarini kashf qilish. Muammolarning ushbu sinfini hal qilish uchun klasterli tahlil usullari qo'llaniladi.

Muammo namunasi: joriy qiymatlarga asoslangan savdo o'sishini bashorat qilish.

Usullari: regressiya, neyron tarmoqlari, genetik algoritmlar, qaror daraxtlari.

Tasniflash va prognozlash muammolari tahlil qilinadigan ob’yekt yoki tizimni o'rganishga olib keladigan induktiv modellashtirish deb nomlanadigan guruhni tashkil qiladi. Ushbu muammolarni hal qilish jarayonida ma'lumotlar bazasi asosida umumiy model yoki faraz ishlab chiqiladi.

Izoh va tavsif

Misol vazifasi: demografik va xarid tarixi bo'yicha mijozlarni tavsiflash.

Usullari: qarorlar daraxti, qoidalar tizimlari, assotsiatsiya qoidalari, bog'lanish tahlili. Agar mijozning daromadi 50 dan oshsa va uning yoshi 30 yoshdan oshsa, u holda mijozning klassi birinchi o'rinda turadi. Umumiy modelni sharhlash bilan tahlilchi yangi bilimlarga ega bo'ladi. Ob’yektlar o'xshashligiga qarab guruhlanadi.

Tushunchalarning o'zaro bog'liqligi. Shunday qilib, biz oldingi ma'ruzada Data Mining usullari va Data Mining bosqichlarida bajarilgan amallarni ko'rib chiqdik. Biz faqat Data Mining-ning asosiy vazifalarini ko'rib chiqdik.

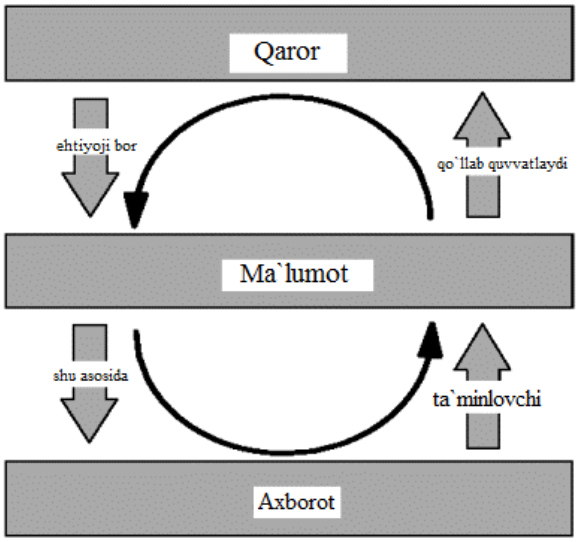
Eslatib o'tamiz, Data Mining-ning asosiy qiymati bu texnologiyaning amaliy yo'nalishi, xom ashyolardan ma'lum bilimlarga, muammoni hal qilishdan tortib, tayyor dasturga o'tish yo'li bo'lib, uning yordamida qarorlar qabul qilinishi mumkin. Data Mining-da birlashtirilgan tushunchalarning xilma-xilligi, shuningdek ushbu texnologiyani qo'llab-quvvatlaydigan usullarning xilma-xilligi, boshlang'ich tahlilchiga mozaikani eslatishi mumkin, uning qismlari bir-biri bilan unchalik bog'liq emas. Qanday qilib biz vazifalar, usullar, harakatlar, shakllar, ilovalar, ma'lumotlar, ma'lumotlar va yechimlarni bog'lashimiz mumkin?

Ikkala oqimni ko'rib chiqamiz:

- **AXBOROT - MA'LUMOT - BILIM VA QO'LLANMALAR**
- **MAVZULAR - ISHLAB CHIQARISH VA QAROR QILISH USULLARI – ILOVALAR**

Ushbu oqimlar "bir xil tanganing ikki tomoni" bo'lib, bitta jarayonning aksidir, natijada bilim va qaror qabul qilish kerak.

Ma'lumotlardan yechimlarga qadar. Birinchi oqimdan boshlaylik. Shaklda – 4.1.1 Qarorlarni qabul qilish jarayonida yuzaga keladigan "axborot", "ma'lumotlar" va "qarorlar" tushunchalari o'rtasidagi o'zaro bog'liqlikni ko'rsatadi.

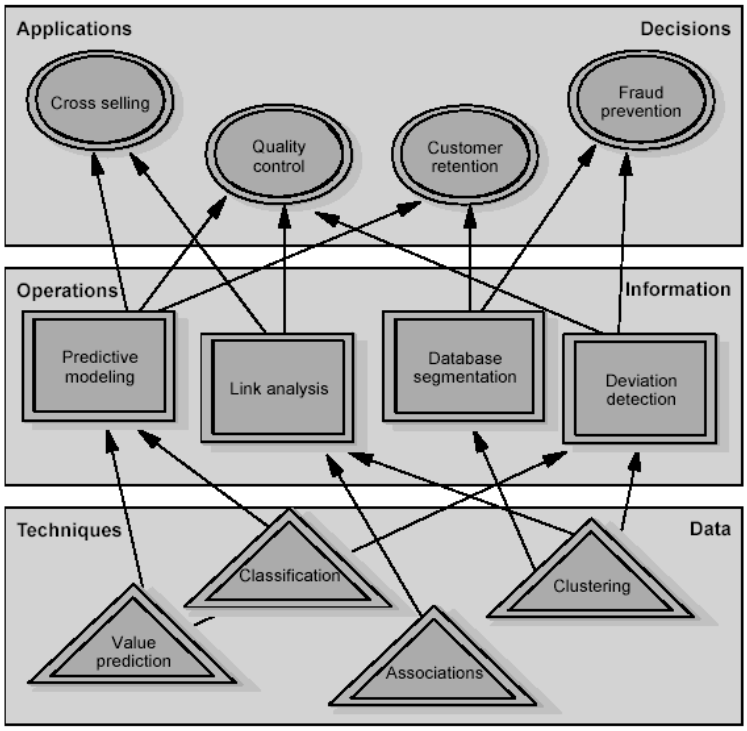


Shakl: 4.1.1. Qarorlar, Axborotlar va ma'lumotlar

Rasmdan ko'rinib turibdiki, bu jarayon tsiklikdir. Qaror qabul qilish ma'lumotlarga asoslangan ma'lumotlarni talab qiladi. Ma'lumotlar qarorlarni qo'llab-quvvatlaydigan ma'lumot va boshqalarni taqdim etadi.

Ko'rib chiqilayotgan tushunchalar axborot piramidasi deb ataladigan ajralmas qism bo'lib, uning bazasida ma'lumotlar, keyingi bosqich - ma'lumotlar, keyin qaror qabul qilinadi, bilim darajasi piramidani to'ldiradi. Axborot piramidasi bo'ylab harakatlanayotganda, ma'lumotlar hajmi echimlarning qiymatiga aylanadi, ya'ni. biznes qiymati. Ma'lumki, Business Intelligence maqsadi ma'lumotlar hajmini biznes qiymatiga aylantirishdir.

Vazifadan tortib dasturgacha. Endi xuddi shu jarayonga boshqa tomondan murojaat qilaylik. 4.1.2-Shaklni ko'rib chiqaylik. Data Mining ta'sir qiladigan barcha darajalarni ko'rsatadi.



Shakl: 4.1.2. Vazifalar, harakatlar, dasturlar

Ta'kidlash joizki, tahlil darajalari (ma'lumotlar, ma'lumotlar, bilimlar) amalda so'nggi yillarda shakllangan ma'lumotlar tahlilining rivojlanish bosqichlariga to'g'ri keladi. Yuqori - dastur darajasi - bu biznes darajasi (agar biz biznes masalasi bilan shug'ullanadigan bo'lsak), unda menejerlar qaror qabul qilishadi. Ko'rsatilgan murojaatlarga misollar: sotish, sifatni boshqarish, mijozlarni ushlab turish.

O'rta - harakatlar darajasi - bu ma'lumot darajasi, aynan shu darajada Data Mining xatti-harakatlari amalga oshiriladi; rasmda quyidagi harakatlar ko'rsatilgan: bashoratli modellashtirish (oldingi ma'ruzada muhokama qilingan), havolani tahlil qilish, ma'lumotlarni segmentatsiyalash va boshqalar.

Eng past ko'rsatkich - bu mavjud bo'lgan ma'lumotlarga nisbatan hal qilinishi kerak bo'lgan ma'lumot konini aniqlash darajasi; rasmda raqamlarni bashorat qilish, tasniflash, klasterlash, birlashtirish muammolari ko'rsatilgan.

Ushbu tushunchalar o'rtasidagi bog'liqlikni ko'rsatadigan jadvalni ko'rib chiqamiz.

3-bosqich	qo'shimchalar	mijozni ushlab turish	bilim	Data Mining natijalari
2-bosqich	harakatlar	*bashoratli modellashtirish*	ma `lumot	tahlil usuli
1-bosqich	vazifalar	Klassifikatsiya	Axborot	so'rovlar

4.1.1-jadval. Ma'lumotlarni qidirish darajalari

Eslatib o'tamiz, tasniflash muammosini hal qilish uchun birinchi bosqichning natijalari (qoida induktsiyasi) yangi ob’yektni ma'lum bir ishonch bilan ma'lum qiymatlarga asoslangan oldindan belgilangan sinflardan biriga tayinlash uchun ishlatiladi. Mijozlarni ushlab turish muammosini ko'rib chiqamiz (firmaning mijozlari ishonchliligini aniqlash).

1-bosqich. Ma'lumotlar - mijozlar bazasi. Mijoz haqida ma'lumotlar mavjud (yosh, jins, kasb, daromad). Mijozlarning ma'lum bir qismi kompaniya mahsulotidan foydalangan holda unga sodiq qolishgan; boshqa mijozlar endi firma mahsulotlarini sotib olmaydilar. Ushbu darajada biz muammoning turini aniqlaymiz - bu tasniflash muammosi.

2-bosqichda biz harakatni aniqlaymiz - bashoratli modellashtirish. Bashoratli modellashtirish yordamida biz ma'lum bir ishonch bilan yangi ob’yektni, yangi mijozni, taniqli sinflardan biriga - doimiy mijozga yoki, ehtimol, bu bir martalik sotib olishga tasniflashimiz mumkin.

3-bosqichda, biz qarorni qabul qilish uchun dasturdan foydalanishimiz mumkin. Bilim olish natijasida firma mijozlarning qaysi biri reklama materiallarini faol ravishda yuborishi kerakligini bilib, reklama xarajatlarini sezilarli darajada kamaytirishi mumkin.

Shunday qilib, bir nechta ma'ruzalar davomida biz "ma'lumotlar", "vazifalar", "usullar", "harakatlar" tushunchalarini yanada mustahkamlaymiz.

2. Axborot va bilim

Endi keling, hali ko'rib chiqilmagan axborot tushunchasiga to'xtalamiz. Ushbu kontseptsiyaning keng tarqalganligiga qaramay, biz uni har doim aniq aniqlay olmaymiz va ma'lumotlar kontseptsiyasidan ajrata olmaymiz. Axborot o'z

mohiyatiga ko'ra ko'p qirrali xususiyatga ega. Insoniyat rivojlanishi bilan, shu jumladan kompyuter texnologiyalari rivojlanishi bilan axborot tobora yangi xususiyatlarga ega bo'lmoqda.

Lug'atga murojaat qilaylik. Axborot (lat.information) –

- biror narsa haqida har qanday xabarlar;
- saqlash, qayta ishlash va uzatish ob’yekti bo'lgan ma'lumotlar (masalan, genetik ma'lumotlar);
- matematikada (kibernetika) - noaniqlikni yo'q qilishning miqdoriy o'lchovi (entropiya), tizimni tashkil etish o'lchovi; axborot nazariyasida - ma'lumot to'plash, uzatish, o'zgartirish va hisoblash bilan bog'liq bo'lgan miqdoriy shakllarni o'rganuvchi kibernetika bo'limi.

Axborot - aniq bir operatsiya ob’yekti bo'lgan voqea, ob’yekt, jarayon va hk haqida oldindan noma'lum bo'lgan har qanday ma'lumot mazmunli talqin qilish uchun. Bu yerda operatsiyalar idrok etish, uzatish, o'zgartirish, saqlash va foydalanishni anglatadi. Axborotni idrok qilish uchun uni talqin qila oladigan, o'zgartira oladigan, ma'lum qoidalarga muvofiqligini aniqlaydigan va hokazo ma'lum idrok etish tizimi kerak. Shunday qilib, axborot tushunchasini faqat ma'lumot manbai va oluvchisi, shuningdek ular orasidagi aloqa kanali mavjud bo'lganda ko'rib chiqish kerak.

Axborot xususiyatlari. Axborotning to'liqligi. Ushbu xususiyat ma'lumotlarning sifatini tavsiflaydi va qarorlarni qabul qilish uchun ma'lumotlarning etarliligini belgilaydi, ya'ni. ma'lumotlarda barcha kerakli ma'lumotlar to'plami bo'lishi kerak.

Misol. "A mahsulotining sotuvi pasayishni boshlaydi" Ushbu ma'lumot to'liq emas, chunki ularning qachon aniq rad etilishi ma'lum emas.

To'liq ma'lumotlarga misol. "Birinchi chorakdan boshlab A mahsulotining sotuvi pasayishni boshlaydi." Ushbu ma'lumot qaror qabul qilish uchun yetarli.

- Axborotning ishonchliligi.

Ma'lumotlar ishonchli yoki ishonchsiz bo'lishi mumkin. Ishonchsiz ma'lumotlarda shovqin paydo bo'ladi va u qanchalik baland bo'lsa, ma'lumotning ishonchliligi shunchalik past bo'ladi.

- Axborotning ahamiyati.

Axborotning qiymati mavhum bo'lolmaydi. Ma'lumotlar foydalanuvchilarning ma'lum bir toifasi uchun foydali va qimmatli bo'lishi kerak.

- Axborotning adolatligi.

Ushbu xususiyat ma'lumotlarning haqiqiy ob’yektiv holatga muvofiqligi darajasini tavsiflaydi. Kerakli ma'lumot to'liq va ishonchli ma'lumotdir.

- Axborotning dolzarbligi.

Ma'lumotlar zamonaviy bo'lishi kerak, ya'ni. eskirmagan. Axborotning bu xossasi ma'lumotlarning hozirgi vaqtga muvofiqligi darajasini tavsiflaydi.

- Axborotning ravshanligi.

Ma'lumotlar odamlar uchun tushunarli bo'lishi kerak.

- Axborotning mavjudligi.

Mavjudlik ma'lum ma'lumotni olish imkoniyatini tavsiflaydi. Axborotning ushbu xususiyatiga ma'lumotlar va tegishli usullarning mavjudligi ta'sir qiladi.

- Axborotning sub'ektivligi.

Axborot sub'ektiv xarakterga ega, u sub'yektni (ma'lumot oluvchisi) idrok etish darajasi bilan belgilanadi.

Axborotga talablar

- Axborotning dinamik xarakteri.

Axborot faqat ma'lumotlar va usullarning o'zaro ta'siri paytida mavjud, ya'ni. axborot jarayoni paytida. Qolgan vaqt ma'lumotlar holatida.

- Amaldagi usullarning mosligi.

Axborot ma'lumotlardan olinadi. Biroq, bir xil ma'lumotlardan foydalanish turli xil ma'lumotlarga olib kelishi mumkin. Bu dastlabki ma'lumotlarni qayta ishlashning tanlangan usullarining etarliligiga bog'liq.

Ma'lumotlar tabiatan ob’yektivdir. Usullar subyektiv, metodlar algoritmlarga asoslangan, subyektiv ravishda tuzilgan va tayyorlangan. Shunday

qilib, ma'lumotlar ob’yektiv ma'lumotlar va sub'ektiv usullarning dialektik o'zaro ta'siri vaqtida paydo bo'ladi va mavjud.

Biznes uchun ma'lumot qaror qabul qilishning boshlang'ich tarkibiy qismidir. Korxona faoliyati va uni boshqarish jarayonida yuzaga keladigan barcha ma'lumotlar ma'lum bir tarzda tasniflanishi mumkin. Qabul qilish manbasiga qarab, ma'lumot ichki va tashqi bo'linadi (masalan, firma tashqarisida yuz beradigan, ammo u bilan bevosita bog'liq bo'lgan hodisalarni tavsiflovchi ma'lumotlar). Shuningdek, ma'lumotni haqiqiy va prognozga ajratish mumkin. Korxona to'g'risidagi faktik ma'lumotlarga muvofiqlikni tavsiflovchi ma'lumotlar kiradi; aniq. Prognoz ma'lumotlari hisoblab chiqilgan yoki bashorat qilingan, shuning uchun uni aniq deb hisoblash mumkin emas, ba'zi bir xatolar bo'lishi mumkin.

Bilim - bu faktlar, naqshlar va evristik qoidalar to'plami, ularning yordamida vazifa hal qilinadi. Shunday qilib, ma'lumotlarning shakllanishi to'plash va uzatish jarayonida sodir bo'ladi, ya'ni. ma'lumotlarni qayta ishlash. Axborotdan bilim qanday olinadi?

Ko'pincha, haqiqiy bilim geterogen ma'lumotlarning taqsimlangan o'zaro bog'liqliklari asosida shakllanadi. Aniq aniqlanmagan natijani olish uchun ma'lumot to'planib, uzatilsa, siz bilimga ega bo'lasiz. Axborotning o'zi sof shaklda ma'nosizdir. Bundan kelib chiqadiki, ma'lumot bu biron bir qo'llaniladigan vositalar yordamida ramzlar shaklida uzatiladigan taktik bilimdir.

Denham Greyning so'zlariga ko'ra, "bilim - bu odamlarning amaliy tajribasi, qobiliyatlari, g'oyalari, sezgi, ishonchi va motivatsiyasi potentsiali bilan birga ma'lumot va ma'lumotlardan mutlaq foydalanish".

Ma'lumot uni ma'lumotdan ajratib turadigan ma'lum xususiyatlarga ega.

Tarkibiylik. Bilimlarni "saralash" kerak.

Kirish va o'rganish qulayligi. Biror kishi uchun bu tez anglash va eslab qolish yoki aksincha eslab qolish qobiliyati; kompyuter bilimi uchun - bilimga kirish vositasi.

Lakonizm. Lakonizm bilimlarni tezda o'zlashtirish va qayta ishlashga imkon beradi va "foydali tarkibiylik koeffitsienti" ni oshiradi. Laconicism ushbu

ro'yxatga kompyuter shovqini va axlat hujjatlari muammosi tufayli ma'lum bo'lgan, chunki bu kompyuter ma'lumotlari - Internet va elektron hujjat aylanishiga xosdir.

Muvofiqlik. Bilim bir-biriga zid bo'lmasligi kerak.

Qayta ishlash tartibi. Undan foydalanish uchun bilim kerak. Bilimning asosiy xususiyatlaridan biri bu uni boshqalarga o'tkazish qobiliyati va unga asoslangan xulosalar chiqarish qobiliyatidir. Buning uchun bilimlarni qayta ishlash protseduralari mavjud bo'lishi kerak. Xulosa chiqarish qobiliyati dastgoh uchun ishlov berish va chiqish protseduralarining mavjudligi va ma'lumotlar tizimlarining bunday ishlov berishga tayyorligini anglatadi. maxsus bilim formatlarining mavjudligi.

"Axborot", "ma'lumotlar", "bilim" tushunchalarini tanlash va taqqoslash

"Axborot", "ma'lumotlar", "bilim" tushunchalari bilan ishonchli ishlash uchun nafaqat ushbu tushunchalarning mohiyatini tushunish, balki ular orasidagi farqni his qilish kerak. Biroq, bu tushunchalarning bitta intuitiv talqini bu erda etarli emas. Yuqoridagi tushunchalar o'rtasidagi farqni tushunish qiyinligi ularning aniq sinonimiyasidadir. Eslatib o'tamiz, Data Mining kontseptsiyasi xuddi shu uchta tushunchadan foydalangan holda rus tiliga tarjima qilingan: ma'lumotlar qidirish, ma'lumot olish, bilimlarni qazib olish.

Birinchidan, oddiy misollardan foydalanib, ushbu atamalarni tushunishga harakat qilaylik.

1. Imtihon topshirgan talabaga axborot kerak.
2. Imtihon topshirgan talabaga ma'lumot kerak.
3. Imtihon topshirgan talaba bilimga muhtoj.

Birinchi variantni ko'rib chiqayotganda - talaba axborotlarga muhtoj - fikr talabaning ma'lumotlarga, masalan, hisob-kitoblarga muhtojligi haqidagi fikrga asoslanadi. Ikkinchi versiyadagi ma'lumotlar tezis yoki darslik bo'lishi mumkin. Ulardan foydalanish natijasida talaba faqat ma'lumotlarga ega bo'ladi, ular ma'lum hollarda bilimga aylanishi mumkin. Uchinchi variant eng mantiqiy ko'rinadi. Axborotlardan farqli o'laroq, ma'lumotlar mantiqiy hisoblanadi.

"Axborot" va "bilim" tushunchalari, falsafiy nuqtai nazardan, nisbatan yaqinda paydo bo'lgan "ma'lumotlar" ga nisbatan yuqori darajadagi tushunchalardir.

"Axborot" tushunchasi axborot tizimidagi jarayonlarning mohiyati bilan bevosita bog'liq, keyin "bilim" tushunchasi ko'proq jarayonlar sifatiga qaratilgan. Bilim qaror qabul qilish jarayoni bilan chambarchas bog'liq.

Farqlarga qaramay, yuqorida aytib o'tilganidek, ko'rib chiqilgan tushunchalar bir-biriga zid emas va o'zaro bog'liq emas. Ular bitta oqimning bir qismidir: uning manbasida ma'lumotlar uzatilishi jarayonida axborot paydo bo'ladi va ma'lum bir sharoitda ma'lumotlardan foydalanish natijasida bilimlar paydo bo'ladi.

Ma'ruzada ta'kidlanganidek, axborot piramidasini siljitish jarayonida ma'lumotlar hajmi bilim qiymatiga aylanadi. Biroq, katta hajmdagi ma'lumotlar umuman ma'no bermaydi va bundan tashqari, bilim olishga kafolat bermaydi. Olingan bilimlarning qiymati ma'lumotlarni qayta ishlash protseduralarining sifati va kuchiga bog'liqligi mavjud. Ma'lumotga aylanib bo'lmaydigan ma'lumotlarning odatiy misoli - bu chet tilidagi matn. Lug'at va tarjimon bo'lmaganda, bu ma'lumotlarning ahamiyati yo'q, u ma'lumotga kira olmaydi. Lug'at yordamida ma'lumotdan bilimga o'tish jarayoni mumkin, ammo ko'p vaqt va mehnat talab qiladi. Tarjimonning ishtirokida ma'lumot haqiqatan ham bilimga aylanadi. Shunday qilib, qimmatbaho bilimlarni olish uchun yaxshi ishlov berish tartib-qoidalari zarur. Ma'lumotdan bilimga o'tish jarayoni ko'p vaqt talab etadi va qimmatga tushadi. Demak, Data Mining texnologiyasi o'zining kuchli va xilma-xil algoritmlariga ega bo'lib, axborot piramidasi yordamida biz haqiqatan ham yuqori sifatli va qimmatli bilimlarga ega bo'lishimiz mumkin.

Nazorat savollari:

1. Data Mining qanday vazifalarni bajaradi?
2. Ma`lumotlarni qidirish deganda nimani tushunasiz?
3. Axborot va bilim tushunchalariga ta`rif bering.
4. Axborotni qidirishning qanday darajlari mavjud?

5-MAVZU

KLASSIFIKATSIYA VA KLASTERIZATSIYA

Reja:

1.Klassifikatsiya.

2.Klasterizatsiya.

**Mashg`ulot maqsadi.** *Ushbu mashg`ulotda ikkita ma'lumot ishlab chiqarish vazifasi - tasniflash va klasterlash batafsil ko'rib chiqilgan. Vazifalarning mohiyati, echish jarayoni, echish usullari, qo'llanilishi tasvirlangan. Ko'rib chiqilgan ikkita muammoning taqqoslanishi keltirilgan.*

Tayanch iboralar: *Ma'lumotlar, tasniflash, klasterlash, kombinatsiyalar, naqsh, moslashuvchanlik, ob'yekt, nazoratsiz o'rganish, nazoratli o'rganish, ikkilik tasniflash, o'zgaruvchan, to'plamlar, bir o'lchovli, ko'p o'lchovli, ko'p o'lchovli tasniflash, ish joylari, ma'lumotlar bazasi, sinf, yorliq, qiymat, atribut jarayon, tasniflagich, yozuv, ma'lumotlar bazasi, o'quv to'plami, to'plam, qurilish, sun'iy neyron tarmoqlari, qo'llab-quvvatlovchi vektor, CBR, genetik algoritm, o'zaro faoliyat tekshirish, aniqlik, namuna olish, baholash, barqarorlik, ishonchlilik, barqarorlik, klaster, taksonomiya, qidiruv, tahlil qilish, klaster, bir-biriga zid kelmaslik, ierarxiya, zichlik, kvantlash, ma'lumot miqdori, bo'linish, SOM.*

1. Klassifikatsiya

Oldingi mavzuda biz Data Mining-ning asosiy vazifalari haqida qisqacha muhokama qildik. Ulardan ikkitasi – Kalassifikatsiya(tasniflash) va klasterlash - biz ushbu mavzuda batafsil ko'rib chiqamiz.

Klassifikatsiya(Tasniflash). Klassifikatsiya eng oddiy va shu bilan birga eng tez-tez hal qilinadigan Data Mining vazifasidir. Tasniflash vazifalarining keng tarqalganligi sababli ushbu kontseptsiyaning mohiyatini aniq tushunish kerak. Bu yerda ba'zi ta'riflar mavjud.

Klassifikatsiya - o'rganilayotgan predmetlar, hodisalar, jarayonlarning jinsi, turlari, turlari bo'yicha, ularni o'rganish qulayligi uchun har qanday muhim belgilar

bo'yicha tizimli ravishda taqsimlash; asl tushunchalarni guruhlash va ularni o'xshashlik darajasini aks ettiruvchi ma'lum bir tartibda joylashtirish.

Klassifikatsiya - bu ba'zi bir printsipga muvofiq buyurtma qilingan ob’yektlar to'plami bo'lib, ular o'xshash tasniflash xususiyatlariga ega (bir yoki bir nechta xususiyatlar), ushbu ob’yektlar o'rtasidagi o'xshashlik yoki farqni aniqlash uchun tanlangan.

Tasniflash quyidagi qoidalarga rioya qilishni talab qiladi:

- har bir bo'linish aktida faqat bitta bazani qo'llash kerak;

- bo'linish mutanosib bo'lishi kerak, ya'ni. aniq kontseptsiyalarning umumiy hajmi bo'linadigan umumiy kontseptsiya hajmiga teng bo'lishi kerak;

- bo'linma a'zolari o'zaro mutlaqo bo'lishi kerak, ularning hajmi bir-biriga zid bo'lmasligi kerak;

- bo'linish ketma-ket bo'lishi kerak.

Farqlanadi:

- tashqi xususiyatga ko'ra tuzilgan va ob'yektlar (jarayonlar, hodisalar) to'plamini kerakli tartibda berishga xizmat qiladigan yordamchi (sun'iy) tasnif;

- ob'yektlar va hodisalarning ichki hamjamiyatini tavsiflovchi muhim xususiyatlarga ko'ra tuzilgan tabiiy tasnif. Bu ilmiy izlanishlarning natijasi va muhim vositasidir, chunki tasniflangan ob’yektlarning naqshlarini o'rganish natijalarini taklif qiladi va birlashtiradi. Tanlangan xususiyatlarga, ularning kombinatsiyasiga va tushunchalarni bo'lish tartibiga qarab tasniflash quyidagicha bo'lishi mumkin:

- **oddiy** - umumiy tushunchani faqat barcha turlar ochilgunga qadar va faqat bir marta bo'lish. Bunday tasniflashning misoli dixotomiya bo'lib, unda faqat ikkita tushuncha bo'linish a'zolari bo'lib, ularning har biri boshqasiga ziddir (ya'ni, printsipga rioya qilinadi: "A va A emas");

- **murakkab** - bitta kontseptsiyani turli asoslar bo'yicha ajratish va bunday sodda bo'linishlarning yaxlitligini birlashtirish. Bunday tasniflashga misol kimyoviy elementlarning davriy jadvali.

Tasniflash deganda biz predmetlarni (kuzatuvlar, voqealar) oldindan ma'lum sinflardan biriga tayinlashni anglatadi.

Klassifikatsiya - bu ma'lum bir guruhning xususiyatlarini aniqlash to'g'risida xulosa chiqarish imkonini beradigan naqsh. Shunday qilib, tasniflash uchun, u yoki bu hodisa yoki ob’yekt tegishli bo'lgan guruhni tavsiflovchi belgilar bo'lishi kerak (odatda, ba'zi qoidalar allaqachon tasniflangan hodisalarni tahlil qilish asosida tuzilgan).

Klassifikatsiya boshqariladigan yoki boshqariladigan ta'lim deb ham yuritiladigan, nazorat qilinadigan o'quv strategiyasini anglatadi.

Klassifikatsiya vazifasi odatda doimiy va / yoki kategoriyali o'zgaruvchilar namunasiga asoslangan kategoriyaga bog'liq o'zgaruvchini (ya'ni, toifaga bog'liq bo'lgan o'zgaruvchini) bashorat qilish deb nomlanadi.

Masalan, firmaning mijozlaridan qaysi biri ma'lum bir mahsulotni potentsial xaridor ekanligini va kim bo'lmaganligini, kompaniyaning xizmatlaridan kim foydalanishini va kim xohlamasligini va hokazolarni taxmin qilishingiz mumkin. Muammoning bu turi ikkilik tasniflash muammolariga tegishli bo'lib, bunda bog'liq o'zgaruvchi faqat ikkita qiymatni olishi mumkin (masalan, ha yoki yo'q, 0 yoki 1). Boshqa klassifikatsiya optsiyasi, agar bog'liq o'zgaruvchi oldindan belgilangan sinflar to'plamidan qiymatlarni olishi mumkin bo'lsa, paydo bo'ladi. Masalan, mijoz qaysi markadagi avtomobilni sotib olishni xohlashini oldindan aytib berish kerak bo'lganda. Ushbu holatlarda ko'plab o'zgaruvchilar bog'liq bo'lgan o'zgaruvchi uchun ko'rib chiqiladi. Klassifikatsiya bir o'lchovli (bitta atribut) va ko'p o'lchovli (ikki yoki undan ortiq atribut) bo'lishi mumkin.

Ko'p o'lchovli klassifikatsiya biologlar tomonidan organizmlarni tasniflash uchun diskriminatsiya masalalarini hal qilish uchun ishlab chiqilgan. Ushbu yo'nalishga bag'ishlangan dastlabki ishlardan biri R. Fisher (1930) ning ishi bo'lib, unda organizmlar fizik parametrlarini o'lchash natijalariga qarab kichik turlarga bo'lingan. Biologiya ko'p o'lchovli tasnif usullarini ishlab chiqish uchun eng mashhur va qulay muhit bo'lib kelgan va shunday bo'lib qolmoqda.

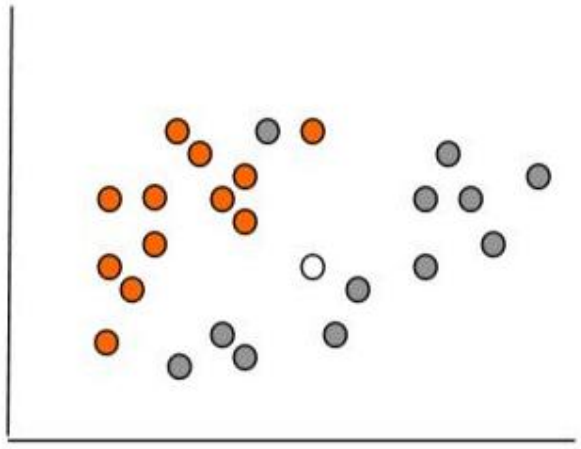
Klassifikatsiya muammosini oddiy misol yordamida ko'rib chiqamiz. Aytaylik, sizning yoshingiz va oylik daromadingiz to'g'risida ma'lumotlarga ega bo'lgan sayyohlik agentligining mijozlari ma'lumotlar bazasi mavjud. Reklama materiallarining ikki turi mavjud: qimmatroq va qulay dam olish va arzonroq, yoshlar ta'tili. Shunga ko'ra, mijozlarning ikkita klassi aniqlanadi: 1-sinf va 2-sinf. Ma'lumotlar bazasi 5.1.1-jadvalda keltirilgan.

Mijoz ID	Yoshi	Foyda	Sinf
1	18	25	1
2	22	100	1
3	30	70	1
4	32	120	1
5	24	15	2
6	25	22	1
7	32	50	2
8	19	45	2
9	22	75	1
10	40	90	2

5.1.1-jadval. Sayyohlik agentligining mijozlar bazasi

Vazifa. Yangi mijoz qaysi sinfga tegishli ekanligini va qaysi turdagi reklama materiallarini yuborishi kerakligini aniqlang.

Aniqlik uchun, keling, bizning ma'lumotlar bazamizni 1-sinfga (apelsin yorlig'i) va 2-sinfga (kul rang yorliq) tegishli ob’yektlar to'plami sifatida ikki o'lchovli o'lchovda (yosh va daromad) taqdim etamiz. Shaklda 5.1.1-da ikkita sinfdagi ob’yektlar ko'rsatilgan.



Shakl: 5.1.1. 2D ichida bir nechta ma'lumotlar bazasi ob’yektlari

Bizning muammoni hal qilish oq yorlig'i bilan rasmda ko'rsatilgan yangi mijoz qaysi sinfga tegishli ekanligini aniqlash bo'ladi.

Klassifikatsiya jarayoni. Klassifikatsiya jarayonining maqsadi predmetli atributlarni kirish sifatida ishlatadigan va bog'liq bo'lgan atributning qiymatini oladigan modelni yaratishdir. Tasniflash jarayoni muayyan mezon bo'yicha ob’yektlar to'plamini sinflarga bo'lishdan iborat.

Tasniflagich(Klassifikator) - bu atributlar vektoriga tegishli predmetning qaysi sinfiga tegishli ekanligini aniqlaydigan muayyan ob’yekt.

Klassifikatsiyani matematik usullardan foydalangan holda amalga oshirish uchun tasniflashning matematik apparati yordamida boshqarilishi mumkin bo'lgan ob’yektning rasmiy tavsifi bo'lishi kerak. Bizning holatda, bunday tavsif ma'lumotlar bazasi. Har bir ob’yekt (ma'lumotlar bazasi yozuvi) ob’yektning ba'zi mulki haqida ma'lumotni olib yuradi. Dastlabki ma'lumotlar to'plami (yoki ma'lumotlar namunasi) ikkita to'plamga bo'lingan: o'quv va sinov.

O'quv to'plami - bu modelni tayyorlash (qurish) uchun ishlatiladigan ma'lumotlarni o'z ichiga olgan to'plam. Ushbu to'plam misollar kirish va chiqish (maqsad) qiymatlarini o'z ichiga oladi. Chiqish qiymatlari modelni o'qitish uchun mo'ljallangan. Sinov to'plamida, shuningdek, misollar kirish va chiqish qiymatlari mavjud. Bu erda, chiqish qiymatlari modelning sog'lig'ini sinash uchun ishlatiladi.

Tasniflash jarayoni ikki bosqichdan iborat: **modelni qurish va undan foydalanish.**

1. **Model tuzilishi:** oldindan belgilangan sinflar tavsifi.

- Har bir misol to'plami bitta oldindan belgilangan sinfga tegishli.

- Ushbu bosqichda model qurilgan o'quv to'plamidan foydalaniladi.

- Olingan model tasniflash qoidalari, qarorlar daraxti yoki matematik formula bilan taqdim etiladi.

1. **Modeldan foydalanish:** yangi yoki noma'lum qiymatlarni tasniflash.

- Modelning to'g'riligini (aniqligini) baholash.

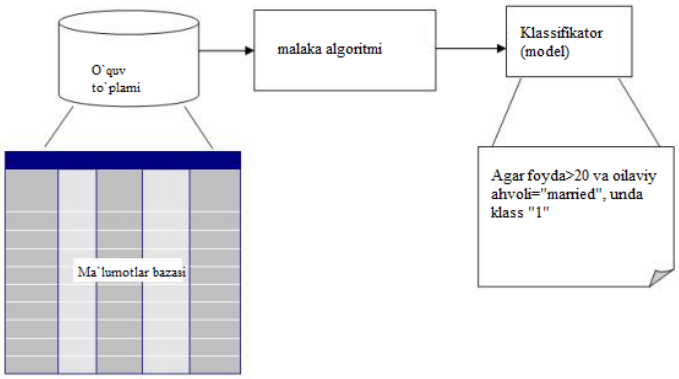
- Sinov voqeasidan ma'lum bo'lgan qiymatlar olingan modelni ishlatish natijalari bilan taqqoslanadi.

- Aniqlik darajasi - bu testlar to'plamidagi to'g'ri tasniflangan misollar foizi.

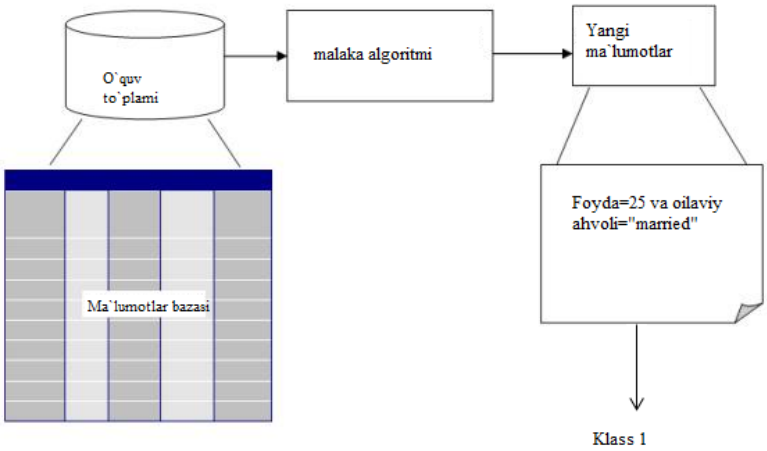
- Sinov to'plami, ya'ni. O'rnatilgan model sinovdan o'tkaziladigan to'plam o'quv to'plamiga bog'liq bo'lmasligi kerak.

- Agar modeldagi aniqlik maqbul bo'lsa, klassi noma'lum bo'lgan yangi misollarni tasniflash uchun modeldan foydalanish mumkin.

Klassifikatsiya jarayoni, aniq namunaviy qurilish va foydalanish sek. 5.1.2. – 5.1.3.



Shakl: 5.1.2. Klassifikatsiya jarayoni. Model qurilishi



Shakl: 5.1.3. Klassifikatsiya jarayoni. Modeldan foydalanish

Klassifikatsiya muammolarini hal qilishda ishlatiladigan usullar

Klassifikatsiya uchun turli xil usullar qo'llaniladi. Ularning asosiylari:

-qaror daraxtlari yordamida tasniflash;

-Basesli (sodda) tasnifi;

-sun'iy neyron tarmoqlari yordamida tasniflash;

-qo'llab-quvvatlovchi vektor tasnifi;

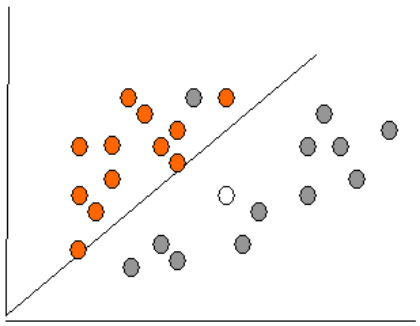
-statistik usullar, xususan, chiziqli regressiya;

-eng yaqin qo'shni usuli yordamida tasniflash;

-CBR usuli bo'yicha tasniflash;

-genetik algoritmlardan foydalangan holda tasniflash.

Klassifikatsiya muammosining ba'zi usullar bilan sxematik echimi (chiziqli regressiya, qaror daraxtlari va neyron tarmoqlari yordamida) sek. 5.1.4 – 5.1.6.



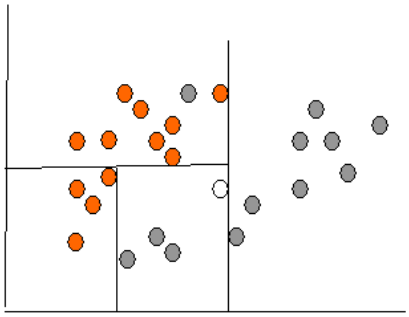
Shakl: 5.1.4. Klassifikatsiya muammosini chiziqli regressiya yordamida hal qilish

if X > 5 then grey

else if Y > 3 then orange

else if X > 2 then grey

else orange



Shakl: 5.1.5. Klassifikatsiya muammosini qaror daraxti usuli bilan hal qilish

Shakl: 5.1.6. Klassifikatsiya muammosini neyron tarmoqlari yordamida hal qilish

Klassifikatsiyaning aniqligi: xatolar darajasini baholash

Klassifikatsiyaning to'g'riligini o'zaro faoliyat tasdiqlash yordamida baholash mumkin. Kross-tekshirish - bu testlar to'plamidagi ma'lumotlar bo'yicha tasniflashning aniqligini baholash protsedurasi bo'lib, uni o'zaro faoliyat tasdiqlash to'plami deb ham atashadi. Sinov to'plamining tasnif aniqligi o'quv to'plamining tasnif aniqligi bilan taqqoslanadi. Agar testlar to'plamini tasniflash o'quv to'plamini tasniflash bilan bir xil natijalarga olib keladigan bo'lsa, ushbu model o'zaro faoliyat tekshiruvdan o'tgan deb hisoblanadi. O'quv va test to'plamlariga bo'linish namunani muayyan nisbatda bo'lish orqali amalga oshiriladi, masalan, o'quv to'plami ma'lumotlarning uchdan ikki qismi va testlar to'plami ma'lumotlarning uchdan biri. Ushbu usul ko'p sonli misollar mavjud bo'lgan namunalar uchun ishlatilishi kerak. Agar namuna kichik hajmga ega bo'lsa, unda mashq qilish va sinov namunalari qisman bir-biriga mos kelishi mumkin bo'lgan maxsus usullardan foydalanish tavsiya etiladi.

Klassifikatsiya usullarini baholash. Usullar quyidagi xususiyatlar asosida baholanishi kerak: tezlik, barqarorlik, tushunarli, ishonchlilik.

Tezlik modelni yaratish va undan foydalanish vaqtini tavsiflaydi.

Mustahkamlik, ya'ni. dastlabki taxminlarning har qanday buzilishlariga qarshilik, shovqinli ma'lumotlar va ma'lumotlarning etishmayotgan qiymatlari bilan ishlash qobiliyatini anglatadi.

Tushuntirish tahlilchiga modelni tushunishga imkon beradi.

Klassifikatsiya qoidalarining xususiyatlari:

- qaror daraxti hajmi;
- tasniflash qoidalarining ixchamligi.

Klassifikatsiya usullarining mustahkamligi ushbu usullarga ma'lumotlar to'plamida shovqin va tashqi narsalar mavjud bo'lganda ishlashga imkon beradi.

2. Klasterizatsiya

Biz nazorat ostidagi o'quv strategiyasi bilan bog'liq klassifikatsiya muammosini ko'rib chiqdik. Ma'ruzaning ushbu qismida biz klasterlash, klaster tushunchalari bilan tanishtiramiz, klasterlash muammosini hal qilishda ishlatiladigan usullar sinflari, klasterlash jarayonining ayrim jihatlari haqida qisqacha to'xtalamiz, shuningdek klaster tahlilini qo'llash misollarini tahlil qilamiz.

Klasterlash vazifasi tasniflash vazifasiga o'xshaydi, uning mantiqiy davomi, ammo farqi shundaki, o'rganilayotgan ma'lumotlar to'plamining sinflari oldindan aniqlanmagan.

Klasterlash avtomatik tasniflash, nazoratsiz o'rganish va taksonomiya bilan sinonimdir.

Klasterlashtirish ob’yektlar to'plamini bir hil guruhlarga (klasterlar yoki sinflar) ajratish uchun mo'ljallangan. Agar namunadagi ma'lumotlar xususiyatlar makonida nuqta sifatida taqdim etilsa, unda klasterlash vazifasi "nuqta kontsentratsiyasi" ta'rifiga tushiriladi.

Klasterlashning maqsadi mavjud tuzilmalarni topishdir. Klasterlash bu tavsiflash protsedurasi bo'lib, u hech qanday statistik xulosalar chiqarmaydi, ammo tahlil ma'lumotlarini va "ma'lumotlar tuzilishini" o'rganishga imkon beradi. Klaster"

tushunchasi noaniq ravishda aniqlanadi: har bir tadqiqotda o'ziga xos "klasterlar" mavjud. Klaster tushunchasi "klaster", "to'plam" deb tarjima qilinadi. Klasterni umumiy xususiyatlarga ega ob’yektlar guruhi sifatida tavsiflash mumkin.

Klaster ikki xususiyatga ega:

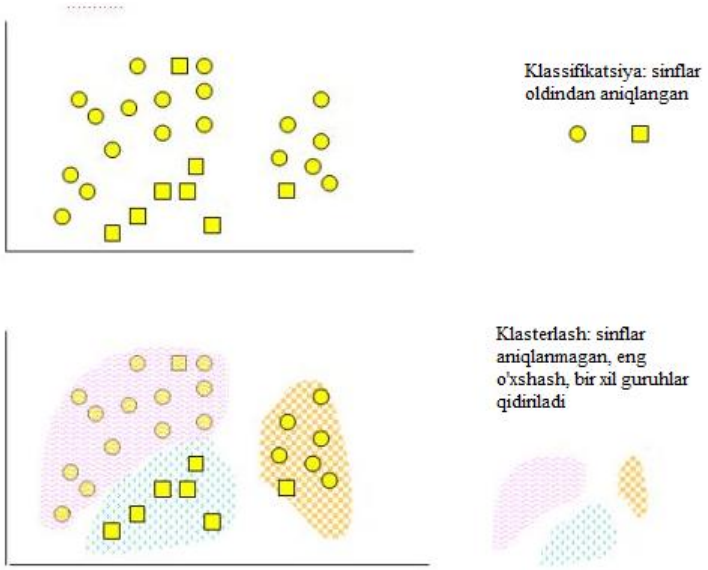
- ichki bir xillik;
- tashqi izolyatsiya.

Ko'pgina muammolarni hal qilishda tahlilchilar savol berishadi - bu ma'lumotlarni vizual tuzilmalarga qanday tashkil qilish, ya'ni. taksonomiyalarni kengaytirish. Dastlab, klasterlash biologiya, antropologiya, psixologiya kabi fanlarda eng katta qo'llanmani oldi. Iqtisodiy muammolarni hal qilish uchun uzoq vaqt davomida iqtisodiy ma'lumotlar va hodisalarning o'ziga xos xususiyatlari tufayli klasterlash kam ishlatilgan.

Xarakteristika	**Klassifikatsiya**	**Klasterizatsiya**
O'rganishning nazorat qilinishi	Nazorat ostidagi ta'lim	Sinovsiz o'rganish
Strategiya	*O'qituvchi bilan o'rganish*	*O'qituvchisiz o'rganish*
Sinf yorlig'i mavjudligi	O'quv to'plamiga kuzatuvga tegishli sinfni ko'rsatuvchi yorliq ilova qilinadi.	O'quv to'plamining sinf yorliqlari noma'lum
Klassifikatsiya uchun asos	Yangi ma'lumotlar o'quv to'plamiga qarab tasniflanadi	Sinflar yoki ma'lumotlar klasterlari mavjudligini aniqlash uchun juda ko'p ma'lumotlar berilgan

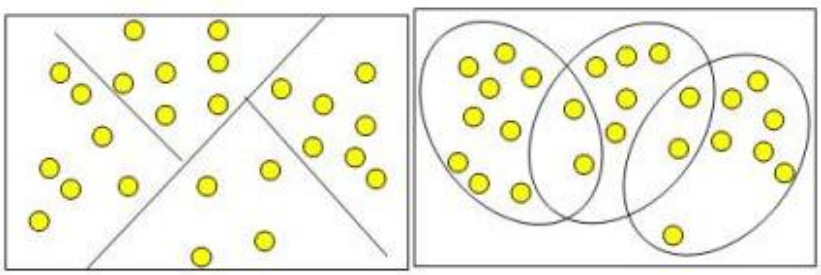
5.2.1-jadvalda klassifikatsiya va klasterlash muammolarining ba'zi parametrlari taqqoslangan.

Shaklda 5.2.1 Klassifikatsiya va klasterlash vazifalarini sxematik ravishda ko'rsatadi.



Shakl: 5.2.1. Klassifikatsiyalash va klasterlash muammolarini taqqoslash

Klasterlar bir-biriga zid bo'lmagan yoki eksklyuziv (bir-biriga mos kelmaydigan, eksklyuziv) va bir-biri bilan qoplangan (bir-biri bilan) bo'lishi mumkin. Parchalanish va kesishgan klasterlarning sxematik ko'rinishi shakl. 5.2.2.



Shakl: 5.2.2. Bir-biri bilan kesishmaydigan va bir-biri bilan kesishadigan klasterlar

Shuni ta'kidlash kerakki, klaster tahlilining turli usullarini qo'llash natijasida turli shakllardagi klasterlarni olish mumkin. Masalan, "zanjir" tipidagi klasterlar mumkin, agar klasterlar uzun "zanjirlar", cho'zilgan shakldagi klasterlar va boshqalar bilan ta'minlangan bo'lsa va ba'zi usullar o'zboshimchalik shaklidagi klasterlarni yaratishi mumkin. Turli xil usullar ma'lum o'lchamdagi klasterlarni yaratishga moyil bo'lishi mumkin (masalan, kichik yoki katta), yoki ma'lumotlar to'plamida turli o'lchamdagi klasterlar mavjudligini taxmin qilish mumkin. Ba'zi klasterlarni tahlil qilish usullari, ayniqsa shovqin yoki chiqadiganlarga sezgir, boshqalari esa kamroq sezgir. Turli xil klasterlash usullarini qo'llash natijasida teng bo'lmagan natijalarga erishish mumkin, bu normal holat va u yoki boshqa

algoritmning ishlashining o'ziga xos xususiyati. Klaster usulini tanlashda ushbu xususiyatlarni hisobga olish kerak. Klaster tahlilining barcha xususiyatlari haqida batafsil uning metodlariga bag'ishlangan ma'ruzada muhokama qilinadi. Bugungi kunga kelib yuzdan ortiq turli klasterlash algoritmlari ishlab chiqilgan. Ko'proq ishlatiladigan ba'zi narsalar ma'ruza kursining ikkinchi qismida batafsil yoritiladi.

Klasterlash yondashuvlari haqida qisqacha ma'lumot.

- Ma'lumotni ajratishga asoslangan algoritmlar (algoritmlarni ajratish), shu jumladan takroriy:

- ob’yektlarni klasterlarga bo'lish;
- Klasterlashni yaxshilash uchun ob’yektlarni iterativ ravishda qayta taqsimlash.

- Ierarxik algoritmlari:

- aglomeratsiya: har bir ob’yekt dastlab klaster, klasterlar bo'lib, ular bir-biri bilan bog'lanib, katta klaster hosil qiladi va hokazo.

- Zichlikka asoslangan usullar:

- ob’yektlarning ulanishi asosida;
- shovqinni e'tiborsiz qoldiring, o'zboshimchalik shaklidagi klasterlarni toping.

- Gridga asoslangan usullar:

- ob’yektlarni panjara tuzilmalariga kiritish.

- Model usullari (Modelga asoslangan):

- ma'lumotlarga mos keladigan klasterlarni topish uchun modeldan foydalaning.

Klaster sifatini baholash

Klaster sifatini baholash quyidagi tartibda amalga oshirilishi mumkin:

- qo'lda tekshirish;

- nazorat punktlarini o'rnatish va natijada paydo bo'lgan klasterlarni tekshirish;

- Modelga yangi o'zgaruvchilar qo'shish orqali klasterlash barqarorligini aniqlash;

- turli usullar yordamida klasterlarni yaratish va taqqoslash. Turli xil klasterlash usullari turli xil klasterlarni yaratishi mumkin va bu normaldir. Shu bilan birga, turli xil usullardan foydalangan holda shunga o'xshash klasterlarni yaratish klasterlash to'g'ri ekanligini ko'rsatadi.

Klasterlash jarayoni. Klaster jarayoni tanlangan usulga bog'liq va deyarli har doim iterativdir. Bu kulgili bo'lishi mumkin va turli xil parametrlarni tanlashda tajribani o'z ichiga olishi mumkin, masalan, masofa o'lchovlari, masalan, standartlashtirish o'zgaruvchilari, klasterlar soni va boshqalar. Biroq, tajribalar o'z-o'zidan tugamasligi kerak - axir, klasterlashning asosiy maqsadi o'rganilayotgan ma'lumotlarning tuzilishi haqida mazmunli ma'lumot olishdir. Olingan natijalar hosil bo'lgan klasterlarni aniq tavsiflash uchun ob’yektlarning xususiyatlari va xususiyatlarini keyingi izohlashni, tadqiq qilishni va o'rganishni talab qiladi.

Klaster tahlilini qo'llash. Klaster tahlili turli sohalarda qo'llaniladi. Bu katta miqdordagi ma'lumotlarni tasniflash zarur bo'lganda foydalidir. Xartigan (1975) nashr etilgan klaster tahlilini o'rganish bo'yicha ko'plab tadqiqotlarni ko'rib chiqdi. Shunday qilib, tibbiyotda kasalliklarni klasterlash, kasalliklarni davolash yoki ularning alomatlarini davolash, shuningdek, bemorlarning taksonomiyasi, dorilar va boshqalar qo'llaniladi. Arxeologiyada tosh konstruktsiyalari va qadimiy buyumlar va boshqalarning taksonomiyalari o'rnatiladi. Marketingda bu raqobatchilar va iste'molchilarni segmentlashtirish vazifasi bo'lishi mumkin. Menejmentda klasterlash vazifasi misolida xodimlarni turli guruhlarga bo'lish, iste'molchilar va etkazib beruvchilarni tasniflash, nikoh ro'y beradigan shunga o'xshash ishlab chiqarish holatini aniqlash mumkin. Tibbiyotda alomatlar tasnifi. Sotsiologiyada klasterlashning vazifasi respondentlarni bir hil guruhlarga bo'lishdir.

Marketing tadqiqotlaridagi klaster tahlili. Marketing tadqiqotlarida klasterli tahlil juda keng qo'llaniladi - ham nazariy tadqiqotlarda, ham turli ob’yektlarni guruhlash muammolarini hal qiluvchi marketologlar tomonidan. Bu mijozlar guruhlari, mahsulotlar va boshqalar haqida savollarni hal qiladi. Shunday qilib, marketing tadqiqotlarida klasterli tahlilni qo'llashda eng muhim vazifalardan biri iste'molchilar xatti-harakatlarini tahlil qilishdir, ya'ni har bir guruhdan mijozning

xulq-atvori va uning xatti-harakatlariga ta'sir qiluvchi omillar to'g'risida xaridorlarni bir hil sinflarga guruhlash. Ushbu muammo Claxton, Fry and Portis (1974), Keel and Layton (1981) asarlarida batafsil tasvirlangan. Klaster tahlili hal qila oladigan muhim vazifa - joylashishni aniqlash, ya'ni. bozorda yangi mahsulotni joylashtiradigan joyni aniqlash. Klaster tahlilini qo'llash natijasida xarita tuzilib, unga binoan bozorning turli segmentlarida raqobat darajasini va ushbu segmentga kirish imkoniyati uchun tovarlarning tegishli xususiyatlarini aniqlash mumkin. Bunday xaritani tahlil qilish orqali bozorda yangi, bo'sh joylarni topish mumkin, ularda mavjud mahsulotlarni taklif qilishingiz yoki yangilarini ishlab chiqishingiz mumkin. Klaster tahlili, masalan, kompaniya mijozlarini tahlil qilish uchun ham foydali bo'lishi mumkin. Buning uchun barcha mijozlar klasterlarga guruhlangan va har bir klaster uchun individual siyosat ishlab chiqilgan. Ushbu yondashuv sizga tahlil qilish ob’yektlarini sezilarli darajada kamaytirishga imkon beradi va shu bilan birga mijozlarning har bir guruhiga individual yondoshadi.

Marketing tadqiqotlarida klaster tahlilidan foydalanish amaliyoti

Marketing tadqiqotlarida klaster tahlilidan foydalanish bo'yicha ba'zi taniqli maqolalar. 1971 yilda mijozlarning xohish-istaklarini tavsiflovchi ma'lumotlar asosida mijozlarni qiziqish doirasi bo'yicha segmentatsiya qilish to'g'risida maqola e'lon qilindi. 1974 yilda Sextonning maqolasi nashr etildi, uning maqsadi mahsulot iste'molchilari bo'lgan oilalar guruhlarini aniqlash edi, natijada brendni aniqlash strategiyalari ishlab chiqilgan. Tadqiqot respondentlarning mahsulot va brendlarga bergan reytinglariga asoslandi. 1981 yilda bir qator o'zgaruvchilardan olingan omil yuklamalari asosida yangi avtomobil sotib oluvchilarning xatti-harakatlarini tahlil qiluvchi maqola e'lon qilindi.

Xulosa. Ushbu ma'ruzada biz klassifikatsiya va klasterlash muammolarini batafsil ko'rib chiqdik. Ushbu vazifalarning o'xshashligi ko'rinishiga qaramay, ular turli yo'llar bilan va turli usullardan foydalangan holda hal qilinadi. Vazifalardagi farq birinchi navbatda dastlabki ma'lumotlarda. Data Mining-ning eng oddiy vazifasi bo'lgan tasniflash "boshqariladigan o'rganish" strategiyasiga tegishli, chunki uni echish uchun o'quv namunasi kirish va chiqish (maqsadli) o'zgaruvchilarning

qiymatlarini o'z ichiga olishi kerak. O'z navbatida, klasterlash - bu ma'lumotni ishlab chiqarishni nazorat qilinmaydigan o'rganish strategiyasi bilan bog'liq, ya'ni. o'quv namunasida maqsadli o'zgaruvchilar qiymatining mavjudligini talab qilmaydi. Klassifikatsiya muammosi turli usullar yordamida hal qilinadi, eng sodda - chiziqli regressiya. Usulni tanlash dastlabki ma'lumotlar to'plamini o'rganishga asoslangan bo'lishi kerak. Klaster muammosini hal qilishda eng keng tarqalgan usullar: k-vositalar usuli (faqat raqamli atributlar bilan ishlaydi), ierarxik klaster tahlili (shuningdek, ramziy atributlar bilan ishlaydi), SOM usuli. Klasterlashning murakkabligi uni baholash zarurati hisoblanadi.

Nazorat savollari

1. Klassifikatsiya nima va uning qanday turlari mavjud?
2. Ko'p o'lchovli klassifikatsiya nima?
3. Klassifikatsiya jarayoniga misol keltiring.
4. Klasterizatsiya nima va uning qanday turlari mavjud?
5. Klasterlashning asosiy vazifasi nimadan iborat?

6-MAVZU

DATA MINING QO`LLASH SOXASI. PROGNOZLASH VA VIZUALIZATSIYA

Reja:

1. **Prognozlash vazifasi.**
2. **Prognozlash va vaqt seriyalari.**
3. **Trend, mavsumiylik va tsikl.**
4. **Prognoz turlari.**
5. **Vizualizatsiya vazifasi.**

Mashg`ulot maqsadi: *Mavzuda prognozlash muammosining mohiyati bayon qilinadi. Vaqt seriyasi tushunchasi, uning tarkibiy qismlari, prognoz parametrlari, prognoz turlari. Ma'lumotni vizualizatsiya qilish muammosi qisqacha tavsiflanadi.*

Tayanch iboralar: *Ma'lumotlar konstruktsiyasi, prognozlash, vizualizatsiya, prognoz, ta'rif, o'quv namunasi, sinf, ma'lumot, vaqt ketma-ketligi, seriya, foiz, klasterlash, segmentatsiya, tasodifiy tanlab olish, tahlil, tendentsiya, mavsumiy tarkibiy, tsiklik tarkibiy qism, bilim, shovqin , prognoz davri, prognoz gorizonti, prognoz oralig'i, prognoz aniqligi, bashoratli model, asboblar to'plami, CHI, SIGGRAPH, IEEE, vizualizatsiya, VA, kompyuter grafikasi, ma'lumotlar grafik tasviri, taqdimot*

Data Mining-ning eng keng tarqalgan va talab qilinadigan vazifalarini ko'rib chiqishda davom etamiz. Ushbu mavzuda biz prognozlash va vizualizatsiya muammolariga to'xtalamiz.

1. Prognozlash vazifasi

Prognozlash vazifalari inson faoliyatining fan, iqtisodiyot, ishlab chiqarish va boshqa ko'plab sohalarida echimi topiladi. Prognozlash har ikkala iqtisodiy sub'ektni va umuman iqtisodiyotni boshqarishni tashkil qilishning muhim elementi hisoblanadi. Prognozlash usullarining rivojlanishi bevosita axborot texnologiyalarining rivojlanishi bilan, xususan, saqlanadigan ma'lumotlar hajmining

ko'payishi va Data Mining vositalarida amalga oshirilayotgan prognozlash usullari va algoritmlarining murakkablashishi bilan bog'liq. Ehtimol, bashorat qilish vazifasi Data Mining-ning eng qiyin vazifalaridan biri deb hisoblanishi mumkin, bu dastlabki ma'lumotlar to'plamini va tahlil qilish uchun mos usullarni sinchkovlik bilan o'rganishni talab qiladi.

Prognozlash (yunoncha Prognosis), so'zning keng ma'nosida, kelajakni oldindan sezuvchi aks sifatida belgilanadi. Prognozlashning maqsadi kelajakdagi voqealarni bashorat qilishdir.

Prognozlash (forecasting) - bu Data Mining-ning vazifalaridan biri va shu bilan birga qarorlarni qabul qilishda muhim jihatlardan biridir.

Prognozlash (prognostics) - bashorat qilish nazariyasi va amaliyoti.

Prognozlash ma'lum bir ob’yekt yoki hodisa dinamikasidagi tendentsiyalarni retrospektiv ma'lumotlar asosida aniqlashga qaratilgan, ya'ni. uning o'tgan va hozirgi holatini tahlil qilish. Shunday qilib, prognozlash muammosini hal qilish uchun ba'zi ma'lumotlarga ega bo'lish kerak.

Prognozlash bu bog'liq va mustaqil o'zgaruvchilar o'rtasida funktsional aloqani o'rnatishdir.

Prognozlash inson faoliyatining ko'plab sohalarida keng tarqalgan va talab qilinadigan vazifadir. Prognozlash natijasida noto'g'ri, asossiz yoki sub'ektiv qarorlarni qabul qilish xavfi kamayadi. Uning vazifalariga misollar: pul oqimlarini bashorat qilish, qishloq xo'jaligi mahsuldorligini bashorat qilish, korxonaning moliyaviy barqarorligini prognoz qilish. Marketingda odatiy vazifa bozorni bashorat qilishdir. Ushbu muammoni hal qilish natijasida ma'lum bir bozor konyunkturasini rivojlantirish istiqbollari, kelgusi davrlar uchun bozor sharoitlarining o'zgarishi baholanadi, bozor tendentsiyalari (tarkibiy o'zgarishlar, mijozlar ehtiyojlari, narxlarning o'zgarishi) aniqlanadi. Odatda ushbu sohada quyidagi amaliy vazifalar hal qilinadi:

- tovarlarni sotish prognozi (masalan, inventarizatsiya stavkasini aniqlash uchun);
- bir-biriga ta'sir ko'rsatadigan mahsulotlarning sotilishini prognoz qilish;

- tashqi omillarga qarab sotish prognozi.

Iqtisodiy va moliyaviy sohalarga qo'shimcha ravishda, prognozlash vazifalari turli sohalarda qo'yilgan: tibbiyot, farmakologiya; siyosiy bashorat qilish endi ommalashmoqda. Umuman olganda, prognozlash muammosini hal qilish quyidagi pastki qismlarni yechishga qisqartiriladi:

- bashorat qilish modelini tanlash;
- tuzilgan prognozning to'g'riligi va aniqligini tahlil qilish.

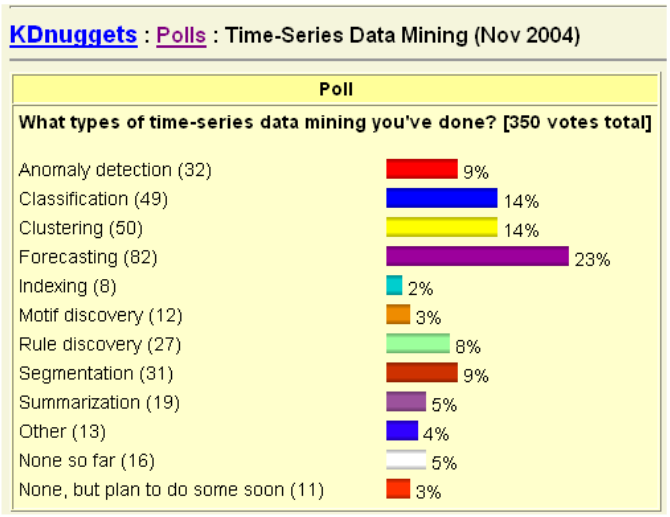
Prognozlash va tasniflash vazifalarini taqqoslash. Oldingi ma'ruzada biz tasniflash muammosini ko'rib chiqdik. Prognozlash tasniflash vazifasiga o'xshaydi.

Tasniflash va prognozlash muammolarini hal qilish uchun ko'pgina ma'lumot qazib olish usullari qo'llaniladi. Bular, masalan, chiziqli regressiya, neyron tarmoqlar, qaror daraxtlari (ba'zan ularni bashorat qilish va tasniflash daraxtlari deb atashadi). Tasniflash va prognozlash muammolari o'xshash va farqlarga ega. Xo'sh, prognozlash va tasniflash muammolari o'rtasida qanday o'xshashliklar mavjud? Ikkala muammo ikki bosqichli jarayondan foydalanib, o'quv to'plamiga asoslangan modelni yaratish va unga bog'liq o'zgaruvchining noma'lum qiymatlarini bashorat qilish uchun foydalanadi. Tasniflash va prognozlash vazifalari o'rtasidagi farq shundaki, birinchi vazifada bog'liq o'zgaruvchining klassi bashorat qilinadi, ikkinchisida esa bog'liq bo'lgan o'zgaruvchan, etishmayotgan yoki noma'lum (kelajakka bog'liq) ning raqamli qiymatlari. Oldingi ma'ruzada muhokama qilingan sayohat agentligi misoliga qaytsak, mijozning sinfini aniqlash tasniflash muammosiga yechim bo'ladi, deb aytishimiz mumkin va kelgusi yilda ushbu mijoz keltiradigan daromadni bashorat qilish prognozlash muammosining yechimi bo'ladi.

2. Prognozlash va vaqt seriyalari

Prognozlash uchun asos vaqt bazalari shaklida saqlanadigan tarixiy ma'lumotlardir. Time-Series Data Mining tushunchasi mavjud. Vaqt ketma-ketligi ko'rinishidagi retrospektiv ma'lumotlarga asoslanib, turli xil Data Mining vazifalarini hal qilish mumkin. 6.2.1.-Shaklda Data Mining vaqt seriyalari bo'yicha so'rov natijalari keltirilgan. Ko'rinib turibdiki, prognozlash echilayotgan vazifalar

orasida eng katta foizni (23%) egallaydi. Keyin tasniflash va klasterlash (har biri 14%), segmentatsiya va anomaliyani aniqlash (har biri 9%), qoidalarni aniqlash (8%). Boshqa vazifalar har birida 6% dan kamroqni tashkil qiladi.



Shakl: 6.2.1. Data Mining vaqti ketma-ketligi

Biroq, prognoz tushunchasiga e'tibor qaratish uchun biz vaqtni faqat prognozlash muammosini hal qilish doirasida ko'rib chiqamiz. Bu yerda vaqt ketma-ketligi va oddiy izlanishlar orasidagi ikkita tub farq bor:

- Vaqt seriyasining a'zolari, tasodifiy tanlanganlar a'zolaridan farqli o'laroq, statistik jihatdan mustaqil emaslar.
- Vaqt seriyalari a'zolari teng taqsimlanmagan.

Vaqt ketma-ketligi - bu xususiyatning tasodifiy bo'lmagan vaqtlarda buyurtma qilingan qiymatlar ketma-ketligi. Vaqt ketma-ketligini tahlil qilish va tasodifiy namunalarni tahlil qilish o'rtasidagi farq kuzatuvlar va ularning xronologik tartiblari o'rtasidagi vaqt oralig'ining tengligini taxmin qilishdir. Bu erda kuzatuvlar vaqti muhim rol o'ynaydi, shu bilan birga tasodifiy tanlovni tahlil qilishda bu muhim emas. Vaqt seriyalarining odatiy misoli - bu birja savdolari ma'lumotlari. Korxonaning turli xil ma'lumotlar bazalarida to'plangan ma'lumotlar xronologik tartibda tuzilgan va ketma-ket keladigan nuqtalarda ishlab chiqarilgan bo'lsa, vaqt qatori hisoblanadi. Vaqt seriyasini tahlil qilish quyidagi maqsadlarda o'tkaziladi:

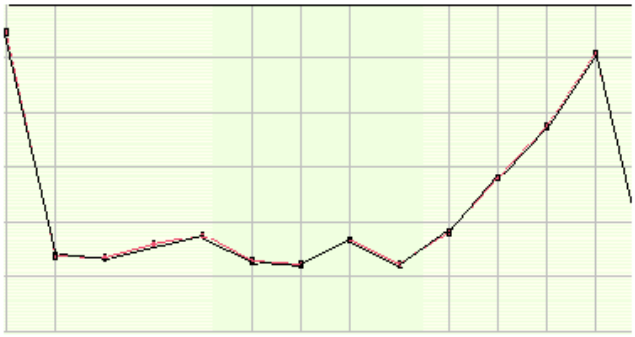
- seriyaning xususiyatini aniqlash;
- seriyaning kelajakdagi qiymatlarini bashorat qilish.

Vaqt ketma-ketligining tuzilishi va naqshini aniqlash jarayonida quyidagilar aniqlanadi: shovqin va chiqindilar, tendentsiya, mavsumiy komponent, tsiklik komponent. Vaqt ketma-ketligining xususiyatini aniqlash ma'lumotlarning o'ziga xos "razvedkasi" sifatida ishlatilishi mumkin. Mavsumiy tarkibiy qism mavjudligi to'g'risida tahlilchining ma'lumotlari, masalan, prognozni tuzishda ishtirok etadigan namunaviy yozuvlar sonini aniqlash uchun zarurdir.

3. Trend, mavsumiylik va tsikl

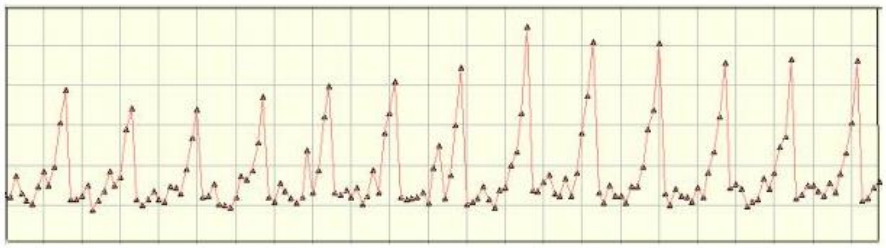
Vaqt seriyasining asosiy tarkibiy qismlari tendentsiya va mavsumiy tarkibiy qismdir. Ushbu turkumlarning tarkibiy qismlari tendentsiyani yoki mavsumiy tarkibiy qismni namoyish qilishi mumkin. Vaqt o'zgarishi mumkin bo'lgan vaqt ketma-ketligining tizimli tarkibiy qismi. **Trend** - bu vaqt ketma-ketligiga ta'sir qiladigan umumiy yoki uzoq muddatli tendentsiyalar ta'siri ostida shakllanadigan tasodifiy bo'lmagan funktsiya. Masalan, o'rganilayotgan bozorning o'sish omili tendentsiyaga misol bo'lishi mumkin. Vaqt seriyasidagi tendentsiyalarni aniqlashning avtomatik usuli yo'q. Ammo vaqt ketma-ketligi monotonik tendentsiyani o'z ichiga olgan bo'lsa (ya'ni, uning doimiy ravishda ko'payishi yoki barqaror pasayishi qayd etilsa), ko'p hollarda vaqt ketma-ketligini tahlil qilish qiyin emas. Prognozlash muammolarining turli xil formulalari mavjud bo'lib, ularni ikki guruhga bo'lish mumkin: bitta seriyali bashorat qilish va ko'p seriyali bashorat qilish yoki o'zaro ta'sir ko'rsatuvchi, seriya. Bir qatorli seriyalarni prognoz qilish guruhi boshqa o'zgaruvchilar va omillarning ta'sirini hisobga olmagan holda faqat ushbu o'zgaruvchiga retrospektiv ma'lumotlar asosida bitta o'zgaruvchining prognozini tuzish vazifalarini o'z ichiga oladi. Ko'p seriyali yoki o'zaro ta'sir ko'rsatadigan bashorat qilish guruhi bir yoki bir nechta o'zgaruvchiga o'zaro ta'sir qiluvchi omillarni hisobga olish zarur bo'lgan tahlil vazifalarini o'z ichiga oladi. Bir seriyali va ko'p seriyali sinflarga bo'lishdan tashqari, seriyalar mavsumiy va mavsumiy emas. Oxirgi bo'linish vaqt seriyasida mavsumiylik kabi tarkibiy qismning mavjudligini yoki yo'qligini anglatadi, ya'ni. mavsumiy tarkibiy qismni kiritish. Vaqt ketma-ketligining mavsumiy tarkibiy qismi vaqt ketma-ketligining davriy takrorlanadigan tarkibiy qismidir. Mavsumiylik xususiyati, vaqtning taxminan teng

vaqt oralig'ida, bog'liq bo'lgan o'zgaruvchining harakatini tavsiflovchi egri shakli uning xarakterli shaklini takrorlashini anglatadi. Mavsumiylik bashorat qilish uchun ishlatiladigan tarixiy ma'lumotlar miqdorini aniqlashda muhimdir. Keling, oddiy misolni ko'rib chiqaylik. 6.3.1. - Shaklda "X mahsulotining sotilishi" o'zgaruvchisining bir oylik davridagi harakatlarini aks ettiruvchi seriyaning parchasi. Rasmda ko'rsatilgan egri chiziqni o'rganayotganda, analitik muntazam ravishda egri shaklining takrorlanishi haqida taxmin qila olmaydi.



Shakl: 6.3.1. Mavsumiy davr uchun vaqt seriyasining bo'limi

Ammo, rasmda ko'rsatilgan uzoqroq seriyani (12 oydan ortiq) ko'rib chiqishda. 6.3.2 - shaklda biz mavsumiy tarkibiy qismning aniq mavjudligini ko'rishingiz mumkin. Shuning uchun, biz bir necha oy davomida ma'lumotlarga qaraganimizda, faqatgina sotish mavsumiyligi haqida gapirishimiz mumkin.



Shakl: 6.3.2. 12 ta fasl davrlaridan vaqt qatorining bo'limi

Shunday qilib, prognoz qilish uchun ma'lumotlarni tayyorlash jarayonida tahlilchi tahlil qiladigan seriya mavsumiylik xususiyatiga ega ekanligini aniqlashi kerak. Mavsumiy komponentning mavjudligini aniqlash kirish ma'lumotlari vakillik xususiyatiga ega bo'lishi uchun zarurdir. Agar uning tashqi ko'rinishini ko'rib chiqayotganda, vaqti-vaqti bilan egri shaklning takrorlanishi haqida taxmin qilish mumkin bo'lmasa, seriyani mavsumiy emas deb hisoblash mumkin. Ba'zan seriyali egri paydo bo'lishidan mavsumiymi yoki yo'qmi, aytish mumkin emas. Mavsumiy

ko'p qator tushunchasi mavjud. Unda har bir satrda bog'liq (maqsadli) o'zgaruvchiga ta'sir etuvchi omillar harakati tavsiflanadi. Bunday seriyalarga misol sifatida bir nechta mavsumiy mahsulotlarning sotuv seriyasini keltirish mumkin. Bunday hollarda ma'lumotlarni to'plash va prognozlash muammosini hal qilish uchun omillarni tanlashda shuni yodda tutish kerakki, tovarlarni sotish hajmining bir-biriga ta'siri mavsumiylik omilining ta'siridan ancha past. Seriyaning mavsumiy tarkibiy qismi va tabiat fasllari haqidagi tushunchalarni chalkashtirmaslik kerak. Ularning tovushlarining o'xshashligiga qaramay, bu tushunchalar boshqacha. Masalan, yozda muzqaymoq sotish boshqa mavsumlarga qaraganda ancha yuqori, ammo bu ushbu mahsulotga bo'lgan talab tendentsiyasidir. Ko'pincha tendentsiya va mavsumiylik bir vaqtning o'zida bir qator ketma-ketlikda mavjud.

Misol. Firmaning foydasi bir necha yillardan beri o'sib bormoqda (ya'ni vaqt qatorida tendentsiya mavjud); seriyasida shuningdek mavsumiy tarkibiy qism mavjud.

Siklik va mavsumiy tarkibiy qismlar o'rtasidagi farqlar:

1. sikl vaqtlari odatda bir mavsumiy davrdan uzunroq;

2. sikllar, mavsumiy davrlardan farqli o'laroq, ma'lum bir vaqtga ega emas.

Har qanday o'zgarishlarni amalga oshirayotganda, vaqt ketma-ketligining tabiatini tushunish osonroq bo'ladi, masalan, tendentsiyani olib tashlash va ketma-ketlikni tekislash.

Prognoz qilishni boshlashdan oldin, siz quyidagi savollarga javob berishingiz kerak:

1. Nimani taxmin qilish kerak?

2. Vaqtinchalik elementlar (parametrlar) nima?

3. Prognoz qanchalik aniq?

Birinchi savolga javob berganda, biz taxmin qilinadigan o'zgaruvchilarni aniqlaymiz. Bu, masalan, keyingi chorakda ma'lum bir mahsulot turini ishlab chiqarish darajasi, ushbu mahsulotni sotish hajmining prognozi va boshqalar bo'lishi mumkin. O'zgaruvchilarni tanlashda tarixiy ma'lumotlarning mavjudligi, qaror qabul qiluvchilarning afzalliklari, Data Miningning yakuniy narxini hisobga olish

kerak. Ko'pincha, prognozlash muammolarini hal qilganda, o'zgaruvchining o'zi emas, balki uning qiymatlari o'zgarishini oldindan aytib berish kerak bo'ladi. Prognozlash muammosini hal qilishda ikkinchi savol quyidagi parametrlarni aniqlashdir:

- Prognoz qilish davri;
- Prognoz qilish gorizonti;
- Prognoz qilish oralig'i.

Prognoz davri prognoz qilinadigan vaqtning asosiy birligidir.

Masalan, biz bir oyda kompaniyaning daromadini bilishni xohlaymiz. Ushbu vazifani bashorat qilish muddati - bir oy.

Prognoz gorizonti - kelajakda prognoz qamrab oladigan davrlar soni.

Agar biz 12 oylik prognozni har oy uchun ma'lumot bilan oldindan bilishni istasak, unda ushbu muammoning prognoz davri - bir oy, prognoz gorizonti - 12 oy.

Bashorat qilish oralig'i - yangi taxmin qilinadigan chastota

Bashorat qilish oralig'i taxmin qilish davri bilan bir xil bo'lishi mumkin. Prognoz parametrlarini tanlash bo'yicha tavsiyalar. Parametrlarni tanlashda bashorat qilish ufqining ushbu prognoz asosida qabul qilingan qarorni amalga oshirish vaqtidan kam bo'lmasligi kerakligini hisobga olish kerak. Shundagina bashorat qilish mantiqiy bo'ladi. Prognoz gorizontining oshishi bilan, prognozning aniqligi, odatda, pasayadi va ufqning pasayishi bilan ortadi. Prognozni amalga oshirish uchun zarur bo'lgan vaqtni qisqartirish va natijada prognozlash xatolarini ufqini kamaytirish orqali prognozlash sifatini yaxshilashimiz mumkin. Prognoz qilish oralig'ini tanlashda ikkita xavfni tanlash kerak: tahlil qilinayotgan jarayonning o'zgarishini o'z vaqtida aniqlamaslik va prognozning yuqori xarajati. Uzoq bashorat qilish oralig'i bilan ushbu jarayonda yuzaga kelgan o'zgarishlarni aniqlamaslik xavfi mavjud, qisqa vaqt ichida prognoz xarajatlari oshadi. Vaqt oralig'ini tanlashda tahlil qilinadigan jarayonning barqarorligini va prognozlash narxini ham hisobga olish kerak.

Pognozning aniqligi. Muayyan muammoni hal qilish uchun zarur bo'lgan prognoz aniqligi prognoz tizimiga katta ta'sir ko'rsatadi. Prognoz xatosi

ishlatilayotgan prognoz tizimiga bog'liq. Bunday tizim qancha ko'p resurslarga ega bo'lsa, shunchalik aniq prognozni olish ehtimoli ko'proq. Biroq, bashorat qilish qaror qabul qilishda xavfni butunlay yo'q qila olmaydi. Shuning uchun mumkin bo'lgan prognoz xatosi har doim hisobga olinadi. Prognozning aniqligi prognoz xatosi bilan tavsiflanadi. Xatolarning eng keng tarqalgan turlari:

- O'rtacha xato (O`X). Bu har bir qadamda xatolarning o'rtacha sonini hisoblash bilan hisoblanadi. Ushbu turdagi xatoning kamchiliklari shundaki, ijobiy va salbiy xatolar bir-birini bekor qiladi.
- O'rtacha mutlaq xato (O`MX). Mutlaq xatolarning o'rtacha qiymati sifatida hisoblanadi. Agar u nol bo'lsa, unda bizda mukammal prognoz mavjud. Kvadratlarning o'rtacha o'rtacha xatosi bilan taqqoslaganda, bu o'lchov tashqi sotuvchiga "juda ko'p ahamiyat bermaydi".
- Kvadrat xatolar yig'indisi (KXY), o'rtacha xato. Kvadratlardagi xatolar yig'indisi (yoki o'rtacha) sifatida hisoblanadi. Bu eng tez-tez ishlatiladigan prognoz aniqligini baholash.
- Nisbiy xato (NX). Oldingi o'lchovlarda haqiqiy xato qiymatlari ishlatilgan. Nisbiy xato mos keladigan sifatni nisbiy xatolar nuqtai nazaridan ifoda etadi.

4. Prognoz turlari

Prognoz qisqa muddatli, o'rta muddatli va uzoq muddatli bo'lishi mumkin.

Qisqa muddatli prognoz - bu bir necha qadam oldinroq bo'lgan prognoz, ya'ni. kuzatuv hajmining 3% dan ko'p bo'lmagan yoki 1-3 qadam oldin bashorat qilingan.

O'rta muddatli prognoz - bu kuzatuvlar hajmining 3-5 foizini tashkil qiladigan bashorat, ammo oldinga 7-12 qadamdan oshmaydi; Shuningdek, ushbu turdagi prognoz mavsumiy tsiklning bir yoki yarmini bashorat qilishni anglatadi. Statistik usullar qisqa va o'rta muddatli prognozlarni tuzishda juda mos keladi.

Uzoq muddatli prognoz - bu kuzatuv hajmining 5% dan ko'prog'ini bashorat qilish. Ushbu turdagi prognozlarni tuzishda statistika usullaridan deyarli foydalanilmaydi, bashoratni shunchaki "tuzish" mumkin bo'lgan juda "yaxshi"

holatlar bundan mustasno. Hozircha biz bashorat qilish usullarini, qaror qabul qilish jarayoni bilan bog'liq ravishda yoki boshqa usullarni ko'rib chiqdik. Prognozlashda e'tiborga olish kerak bo'lgan boshqa omillar mavjud.

Maqsad 1. Ma'lumki, tahlil qilinadigan jarayon vaqt ichida nisbatan barqaror bo'lib, o'zgarishlar asta-sekin sodir bo'ladi, jarayon tashqi omillarga bog'liq emas.

Maqsad 2. Tahlil qilinayotgan jarayon beqaror va tashqi omillarga juda bog'liq.

Birinchi muammoning echimi katta miqdordagi tarixiy ma'lumotlardan foydalanishga qaratilishi kerak. Ikkinchi muammoni hal qilishda predmet sohasidagi mutaxassisning, ekspertning barcha zarur tashqi omillarni bashorat qilinadigan modelda aks ettirishi, shuningdek ushbu omillar bo'yicha ma'lumotlarni to'plash uchun vaqt ajratishi uchun alohida e'tibor berilishi kerak (tashqi ma'lumotlarni to'plash ko'pincha ichki ma'lumot to'plashdan ko'ra ancha qiyinroq). tizimlari). Prognozlash asosida amalga oshiriladigan ma'lumotlarning mavjudligi prognoz modelini yaratishda muhim omil hisoblanadi. Yaxshi prognoz qilish uchun ma'lumotlar ishonchli, aniq va ishonchli bo'lishi kerak.

Prognozlash usullari. Ma'lumotni qazib olish usullari, ularning yordamida prognozlash muammolari hal qilinadi, kursning ikkinchi qismida muhokama qilinadi. Prognozlashda ishlatiladigan keng tarqalgan ma'lumotlar qidirish usullari orasida biz neyron tarmoqlari va chiziqli regressiyani ta'kidlaymiz. Prognozlash usulini tanlash ko'plab omillarga, shu jumladan prognoz parametrlariga bog'liq. Usulni tanlash tarixiy ma'lumotlar to'plamining barcha o'ziga xos xususiyatlarini va u qurilayotgan maqsadlarni hisobga olgan holda amalga oshirilishi kerak. Prognozlashda ishlatiladigan Data Mining dasturi foydalanuvchiga aniq va ishonchli prognozlashni ta'minlashi kerak. Shu bilan birga, bunday prognozni olish nafaqat dasturiy ta'minot va uning asosidagi usullarga, balki boshqa omillarga, jumladan, dastlabki ma'lumotlarning to'liqligi va ishonchliligi, ularni to'ldirishning o'z vaqtida va tezkorligi va foydalanuvchining malakasiga bog'liq.

5.Vizualizatsiya vazifasi

Vizualizatsiya - bu hisob-kitoblarning yakuniy natijasini ko'rish, hisoblash jarayonini nazorat qilishni tashkil qilish va hatto keyingi harakatning eng oqilona yo'nalishini aniqlash uchun dastlabki ma'lumotlarga qaytish imkonini beradigan vositalar to'plami. Vizualizatsiya vazifasini konferentsiya materiallari, masalan, CHI va ACM-SIGGraph, shuningdek davriy nashrlarda, xususan, "IEEE Trans. Vizualizatsiya va kompyuter grafikasi" jurnalining materiallarida batafsil topish mumkin. Vizualizatsiyadan foydalanish natijasida ma'lumotlarning grafik tasviri yaratiladi. Vizualizatsiyadan foydalanish ma'lumotlarni tahlil qilish jarayonida anomaliyalar, tuzilmalar, tendentsiyalarni ko'rishga yordam beradi. Prognozlash muammosini ko'rib chiqayotganda, biz vaqt ketma-ketligining grafik tasvirlanishidan foydalandik va unda mavsumiy tarkibiy qism mavjudligini ko'rdik. Oldingi ma'ruzada biz tasniflash va klasterlash muammolarini ko'rib chiqdik va ob’yektivlarning ikki o'lchovli makonda tarqalishini tasvirlash uchun vizualizatsiyadan ham foydalandik. Aytishimiz mumkinki, vizualizatsiyadan foydalanish yanada tejamkor: tendentsiya chizig'i yoki scatterplot ustidagi nuqtalar to'plami tahlilchiga naqshni tezroq aniqlashga va kerakli echimga erishishga imkon beradi. Shunday qilib, bu erda biz Data Mining-da belgilar emas, balki tasvirlardan foydalanish haqida gaplashamiz.

Vizualizatsiyaning asosiy afzalligi - foydalanuvchilarning maxsus tayyorgarligiga ehtiyojning deyarli yo'qligi. Vizualizatsiya yordamida ma'lumot bilan tanishish juda oson, shunchaki unga qarash kerak. Vizualizatsiyaning eng oddiy turlari ancha oldin paydo bo'lgan bo'lsa-da, undan foydalanish tobora kuchayib bormoqda. Vizualizatsiya nafaqat tahlil usullarini takomillashtirish bilan bog'liq emas - Skott Leybening so'zlariga ko'ra, ba'zi hollarda vizualizatsiya hatto uni almashtirishi mumkin. Ma'lumotni vizualizatsiya qilish quyidagi shakllarda taqdim etilishi mumkin: grafikalar, jadvallar, gistogrammalar, diagrammalar va boshqalar. Qisqacha vizualizatsiya rolini quyidagi xususiyatlar bilan tavsiflash mumkin:

- interfaol va izchil tavsiflash bo'yicha yordam-yordam;
- natijalarni taqdim etish bo'yicha yordam berish;

- vizual tasvirlarni o'qish va tushunish.

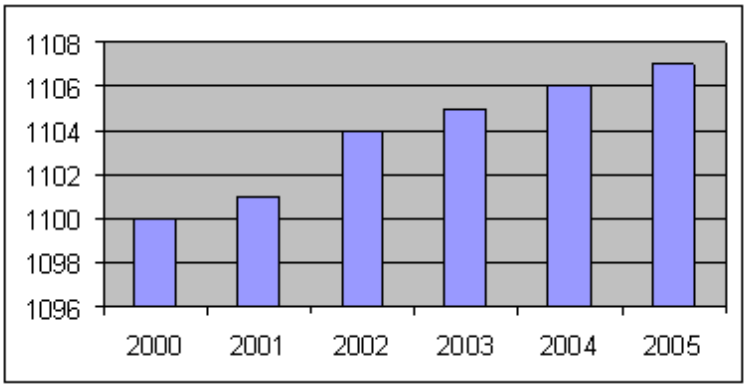
Yomon vizualizatsiya. Natijalarni berish ba'zida foydalanuvchini chalkashtirib yuborishi mumkin. Yomon vizualizatsiya uchun oddiy misol. Aytaylik, bizda "A kompaniyasining foydasi" 2000 yildan 2005 yilgacha bo'lgan davrga ega, 6.5.1-jadvalda jadval shaklida keltirilgan.

Yil	Foyda
2000	1100
2001	1101
2002	1104
2003	1105
2004	1106
2005	1107

6.5.1-jadval. A kompaniyasining foydasi

Keling, ushbu ma'lumotlar uchun Excelda gistogrmmasini tuzamiz.

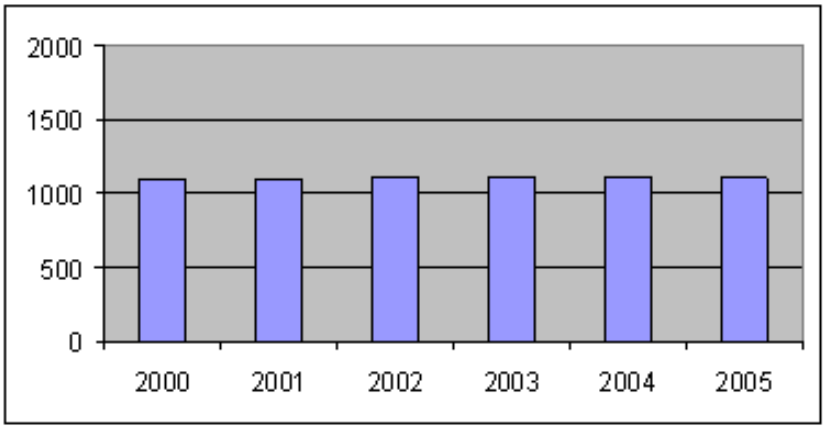
Gistogramma - bu ma'lumotlarning taqsimlanishi ingl. Ushbu ma'lumotlar to'rtburchaklar yoki teng kenglikdagi chiziqlar yordamida ko'rsatiladi, ularning balandligi har bir sinfdagi ma'lumotlar miqdorini ko'rsatadi. Grafikni chizish uchun barcha standart qiymatlardan foydalanib, sekstda ko'rsatilgan gistogrammani olamiz. 6.5.2.



Shakl: 6.5.2. Gistogramma, y o'qining minimal qiymati – 1096

Ushbu raqam A kompaniyasining 2000 yildan 2005 yilgacha daromadlari sezilarli darajada o'sganligini ko'rsatadi. Ammo, agar foyda miqdorini ko'rsatadigan y-o'qiga qarasak, bu o'qning o'qi 1096 qiymatida x o'qini kesib o'tganligini ko'rishimiz mumkin. Aslida y-o'qi 1096 dan 1108 gacha bo'lgan qiymatlar

foydalanuvchini chalg'itmoqda. Y o'qi formati uchun mas'ul bo'lgan parametrlarning qiymatlarini o'zgartirgan holda, rasmda ko'rsatilgan grafikka ega bo'lamiz. 6.5.3.-shakl.



Shakl: 6.5.3. Bar chizmasi, minimal y o'qining qiymati 0 ga teng

0 dan 2000 gacha bo'lgan qiymatlarga ega bo'lgan y o'qi foydalanuvchiga kompaniya foydasining ozgina o'zgarishi haqida to'g'ri ma'lumot beradi. Dastlabki ma'lumotlarning katta o'lchamlari va murakkabligi haqida gap ketganda, vizualizatsiya vositalari keskin kamayishni ta'minlaydi va ehtimol millionlab ma'lumotlar yozuvlarini sodda, tushunish va boshqarishni osonlashtiradi. Bunday namoyishlar ma'lumotlarning vizual yoki grafik namoyishlari deyiladi. Vizualizatsiya Data Mining vositalaridan foydalanib olingan ma'lumotlarni o'rganishda muhim omil sifatida qaralishi mumkin. Bunday holatlarda biz vizual Data Mining haqida gapiramiz. Vizualizatsiya usullari, shu jumladan ma'lumotni bir, ikki, uch o'lchovli va kattaroq o'lchamlarda taqdim etish, shuningdek ma'lumotni namoyish qilishning boshqa usullari, masalan, parallel koordinatalar, "Chernov yuzi", kursning keyingi qismida muhokama qilinadi.

Nazorat savollari:

1. Prognozlash vaqti deganda nimani tushunasiz? Misollar keltiring.
2. Trend, mavsumiylik va tsikl haqida nimalarni bilasiz?
3. Prognozning qanday turlari mavjud?
4. Vizualizatsiya asosiy vazifasi nima?

7-MAVZU

BANK ISHI. SUG`URTA. TELEKOMMUNIKATSIYA. MARKETING. FOND BOZORI. BIOINFORMATIKA. MEDITSINA. FARMASEVTIKA.

Reja:

1. **Bioinformatika Bank ishi.**
2. **Sug`urta.**
3. **Telekommunikatsiya.**
4. **Marketing.**
5. **Fond bozori.**
6. **Bioinformatika**
7. **Meditsina**
8. **Farmasevtika**

Mashg`ulot maqsadi: *Ma'ruzada Data Mining texnologiyasini muvaffaqiyatli qo'llash mumkin bo'lgan inson faoliyatining asosiy yo'nalishlari muhokama qilinadi. Web Mining, Text Mining, Call Mining kabi tushunchalar muhakama qilinadi.*

Kalit so`zlar: *Data Mining, xarajatlar, fond bozorlari, search engine, Web Mining, moliyaviy vositalar, bozor tarkibi, microarray, data analysis, MDA, Web Content Mining, Web Usage Mining, approach, FAQ, Occam, welding, Text Mining, , Call Mining, Audio Mining, Video Mining.*

Avvalgi ma'ruzalarda biz Data Mining vazifalari va usullarini ko'rib chiqdik. Biroq, ushbu texnologiyadan qanday aniq vazifalar va inson hayotining qaysi sohalarida foydalanish mumkinligini o'ylamasak, kirish qismi to'liq bo'lmaydi. Darhol aytish kerakki, Data Mining-dan foydalanish sohasi hech narsa bilan cheklanmaydi - bu ma'lumotlar mavjud bo'lgan hamma joyda. Ushbu ma'ruzada biz Data Mining dasturlarining barcha turlarini ko'rib chiqamiz.

Ushbu sharhning maqsadi dasturning mutlaqo barcha yo'nalishlarini ro'yxatga olish emas, balki Data Mining ishlaydigan va haqiqiy natijalar beradigan sohalar bilan tanishishdir.

Shuni ta'kidlash kerakki, bugungi kunda Data Mining texnologiyasi biznes muammolarini hal qilishda eng keng qo'llaniladi. Ehtimol, sababi aynan shu yo'nalishda Data Mining vositalaridan foydalanishning rentabelligi ba'zi manbalarga ko'ra 1000% gacha bo'lishi mumkin va uni amalga oshirish xarajatlari etarlicha tez to'lanishi mumkin.

Endilikda Data Mining texnologiyasi retrospektiv ma'lumotlar to'plangan inson faoliyatining deyarli barcha sohalarida qo'llaniladi.

1. Bank ishi

Biz Data Mining texnologiyasini qo'llashning to'rtta asosiy yo'nalishini batafsil ko'rib chiqamiz: ilm-fan, biznes, hukumat uchun tadqiqotlar va Internet.

1. Ilmiy tadqiqotlar uchun Data Mining dasturi. Asosiy yo'nalishlari: tibbiyot, biologiya, molekulyar genetika va gen muhandisligi, bioinformatika, astronomiya, amaliy kimyo, giyohvandlikka oid tadqiqotlar va boshqalar.
2. Biznes muammolarini hal qilish uchun Data Mining dasturi. Asosiy yo'nalishlari: bank, moliya, sug'urta, CRM, ishlab chiqarish, telekommunikatsiya, elektron tijorat, marketing, fond bozori va boshqalar.
3. Data Mining dasturini davlat darajasidagi muammolarni hal qilishda qo'llash. Asosiy yo'nalishlar: soliq to'lamaganlarni qidirish; terrorizmga qarshi kurashda anglatadi.
4. Web-muammolarni hal qilish uchun ma'lumotlarni tanlab olishdan foydalanish. Asosiy yo'nalishlar: qidiruv tizimlari, hisoblagichlar va boshqalar.

Biznes muammolarini hal qilish uchun Data Mining dasturi

Bank ishi: Data Minning texnologiyasi bank sohasida bir qator odatiy vazifalarni hal qilishda qo'llaniladi.

"Mijozga qarz berish kerakmi?"

Bank xizmatida Data Mining dasturining klassik namunasi bank mijozining mumkin bo'lgan to'lov qobiliyatini aniqlash muammosini hal qilishdir. Ushbu vazifa, shuningdek, mijozning kredit qobiliyatini tahlil qilish yoki "Mijozga qarz berish kerakmi?"

Data Mining texnologiyasidan foydalanmasdan, muammoni bank muassasasi xodimlari o'zlarining tajribalari, sezgi va qaysi mijoz ishonchli ekanligi to'g'risida sub'ektiv g'oyalariga asoslanib hal qilishadi. Data Minning usullariga asoslangan qarorlarni qo'llab-quvvatlash tizimlari xuddi shunday sxema bo'yicha ishlaydi. Tarixiy (retrospektiv) ma'lumotlarga asoslangan va tasniflash usullaridan foydalangan holda bunday tizimlar ilgari kreditlarni to'lamagan mijozlarni aniqlaydi.

"Mijozga qarz berish kerakmi?" Data Mining usullaridan foydalanish quyidagicha hal qilinadi. Bank mijozlari yig'indisi ikki sinfga bo'linadi (qaytganlar va qaytarmaganlar); kreditni qaytarib bermagan mijozlar guruhi asosida potentsial defoltning asosiy "xususiyatlari" aniqlanadi; yangi mijoz haqida ma'lumot kelganda uning klassi aniqlanadi ("qarzni qaytaradi", "qarzni qaytarmaydi").

Bankning yangi mijozlarini jalb qilish vazifasi.

Data Mining vositalari yordamida "ko'proq foydali" va "kam rentabelli" mijozlarga ajratish mumkin. Mijozlarning eng daromadli segmentini aniqlagandan so'ng, bank aniq mijozlar guruhini jalb qilish uchun yanada faol marketing siyosatini olib borishi mantiqan to'g'ri keladi.

Mijozlarni segmentatsiyalashning boshqa vazifalari.

Data Minning vositalari yordamida mijozlarni turli guruhlarga ajratish orqali bank marketing siyosatini yanada aniqroq va shu sababli samarali qilib, mijozlarning turli guruhlariga aynan kerakli xizmat turlarini taklif qilish imkoniyatiga ega.

Bank likvidligini boshqarish vazifasi. Mijozlarning hisobvaraqlaridagi qoldiqni prognoz qilish. Data Minning usullaridan foydalangan holda, mijozlar hisobvarag'idagi qoldiqlar haqidagi ma'lumotlar bilan vaqt seriyasini prognoz qilish orqali kelajakda ma'lum bir vaqtda hisobvaraqlardagi qoldiq prognozini olishingiz

mumkin. Olingan natijalar bank likvidligini baholash va boshqarish uchun ishlatilishi mumkin.

Kredit kartalar bilan firibgarlik holatlarini aniqlash vazifasi. Shubhali kredit karta operatsiyalarini aniqlash uchun, keyinchalik firibgarlikka aylangan bank operatsiyalarini tahlil qilish natijasida aniqlanadigan "shubhali xatti-harakatlar uslubi" ishlatiladi. Shubhali holatlarni aniqlash uchun ma'lum bir vaqt oralig'ida ketma-ket operatsiyalar to'plamidan foydalaniladi. Agar Data Mining tizimi boshqa operatsiyani shubhali deb hisoblasa, bank xodimi ushbu ma'lumotlarga asoslanib, ma'lum bir karta bilan operatsiyalarni bloklashi mumkin.

2. Sug`urta

Sug'urta biznesi ma'lum bir xavf bilan bog'liq. Bu erda Data Mining yordamida hal qilingan vazifalar bank ishlariga o'xshashdir.

Mijozlarni guruhlarga ajratish natijasida olingan ma'lumotlar mijozlar guruhlarini aniqlash uchun ishlatiladi. Natijada sug'urta kompaniyasi eng katta foyda va eng kam tavakkalchilik bilan mijozlarning ma'lum guruhlariga ma'lum xizmat guruhlarini taklif qilishi mumkin.

3. Telekommunikatsiya

Telekommunikatsiyalarda Data Mining yutuqlari ishonchli mijozlarni jalb qilish uchun ishlaydigan har qanday kompaniyaga xos bo'lgan muammolarni hal qilishda ishlatilishi mumkin - bu mijozlarning sodiqligini aniqlash. Bunday muammolarni hal qilish zarurati telekommunikatsiya bozoridagi qattiq raqobat va mijozlarning doimiy ravishda bir kompaniyadan ikkinchisiga ko'chib o'tishi bilan bog'liq. Ma'lumki, mijozni saqlab qolish uni qaytarishdan ko'ra ancha arzon. Shu sababli, xaridorlarning ayrim guruhlarini aniqlash va ular uchun eng jozibali xizmatlar to'plamini ishlab chiqish zaruriyati tug'iladi. Bu sohada, boshqa ko'plab sohalarda bo'lgani kabi, firibgarlik faktlarini aniqlash muhim vazifadir.

Faoliyatning ko'plab yo'nalishlari uchun xos bo'lgan bunday vazifalardan tashqari, telekommunikatsiya sohasining o'ziga xos xususiyatlari bilan belgilanadigan vazifalar guruhi ham mavjud.

Elektron tijorat. Elektron tijorat sohasida Data Mining tavsiyalar tizimlarini shakllantirish va veb-saytga tashrif buyuruvchilarni tasniflash muammolarini hal qilish uchun ishlatiladi. Ushbu tasnif kompaniyalarga aniq mijozlar guruhlarini aniqlash va aniqlangan mijozlarning qiziqishlari va ehtiyojlariga muvofiq marketing siyosatini olib borish imkoniyatini beradi. Elektron tijorat uchun Data Mining texnologiyasi Web Mining texnologiyasi bilan chambarchas bog'liq.

Sanoat ishlab chiqarishi. Sanoat ishlab chiqarish xususiyatlari va texnologik jarayonlar Data Mining texnologiyasidan turli ishlab chiqarish muammolarini hal qilishda foydalanish imkoniyatlari uchun yaxshi shart-sharoitlarni yaratadi. Texnik jarayon o'z mohiyatiga ko'ra boshqarilishi kerak va uning barcha og'ishlari ilgari ma'lum bo'lgan chegaralar doirasida. Bu yerda odatda Data Mining texnologiyasi oldida turgan vazifalarning ko'pchiligiga xos bo'lmagan ma'lum bir barqarorlik haqida gapirish mumkin.

Data Mining-ning sanoat ishlab chiqarishidagi asosiy vazifalari:

- ishlab chiqarish vaziyatlarini kompleks tizim tahlili;
- ishlab chiqarish vaziyatlarini rivojlanishining qisqa va uzoq muddatli prognozi;
- optimallashtirish echimlari variantlarini ishlab chiqish;
- texnologik jarayonning ba'zi parametrlariga qarab mahsulot sifatini bashorat qilish;
- ishlab chiqarish jarayonlarining rivojlanishining yashirin tendentsiyalari va qonuniyatlarini aniqlash;
- ishlab chiqarish jarayonlarining rivojlanish qonuniyatlarini bashorat qilish;
- yashirin ta'sir etuvchi omillarni aniqlash;
- ishlab chiqarish parametrlari va ta'sir etuvchi omillar o'rtasidagi ilgari noma'lum munosabatlarni aniqlash va aniqlash;
- ishlab chiqarish jarayonlarining o'zaro ta'sir muhitini tahlil qilish va uning xususiyatlarining o'zgarishini bashorat qilish;
- ishlab chiqarish jarayonlarini boshqarish bo'yicha optimallashtirish bo'yicha tavsiyalar ishlab chiqish;

- tahlil natijalarini vizuallashtirish, mumkin bo'lgan amalga oshirishning ishonchliligi va samaradorligini baholash bilan dastlabki hisobotlarni va amalga oshiriladigan echimlar loyihalarini tayyorlash.

4. Marketing

Data Mining marketingda keng qo'llaniladi. Asosiy marketing savollari "Nima sotilmoqda?", "Qanday sotilmoqda?", "Iste'molchi kim?" Tasniflash va klasterlash muammolariga bag'ishlangan ma'ruzada iste'molchilar segmentatsiyasi kabi marketing muammolarini hal qilish uchun klaster tahlilidan foydalanish batafsil bayon etilgan. Marketing muammolarini hal qilishning yana bir keng tarqalgan usullaridan biri bu assotsiatsiya qoidalarini topish usullari va algoritmlari. Bu erda vaqtinchalik naqshlarni qidirish ham muvaffaqiyatli qo'llanilmoqda.

Chakana savdo sohasida, shuningdek marketingda quyidagilar qo'llaniladi:

- assotsiatsiya qoidalarini topish algoritmlari (xaridorlar bir vaqtning o'zida sotib oladigan tez-tez uchrab turadigan tovar to'plamlarini aniqlash uchun). Ushbu qoidalarni aniqlash tovarlarni savdo maydonchalari javonlariga joylashtirishga, tovarlarni sotib olish va ularni omborlarga joylashtirish strategiyasini ishlab chiqishga va boshqalarga yordam beradi.
- vaqt ketma-ketliklaridan foydalanish, masalan, omborda kerakli miqdordagi tovar zaxirasini aniqlash.
- mijozlarning guruhlari yoki toifalarini aniqlash uchun tasniflash va klasterlash usullari, ularning bilimlari tovarlarni muvaffaqiyatli reklama qilishga yordam beradi.

5. Fond bozori

Data Minning texnologiyasi yordamida hal qilinishi mumkin bo'lgan fond bozori muammolari ro'yxati:

- moliyaviy vositalarning kelajakdagi qiymatlarini va ularning o'tgan qiymatlari asosida ko'rsatkichlarni prognoz qilish;
- moliyaviy vositaning tendentsiya prognozi (harakatning kelajakdagi yo'nalishi - o'sish, pasayish, yassi) va uning kuchi (kuchli, o'rtacha kuchli va hk);

- ma'lum xususiyatlar to'plami bo'yicha bozor, sanoat, sektorning klaster tuzilishini aniqlash;
- portfelni dinamik boshqarish;
- o'zgaruvchanlik prognozi;
- xavf-xatarni baholash;
- inqiroz boshlanishini bashorat qilish va uning rivojlanishini bashorat qilish;
- aktivlarni tanlash va boshqalar.

Data Mining texnologiyasi yuqorida tavsiflangan faoliyat yo'nalishlaridan tashqari, ma'lumotlarni tahlil qilish zarurati bo'lgan va ma'lum miqdordagi retrospektiv ma'lumot to'plangan turli xil biznes sohalarida qo'llanilishi mumkin.

CRM(customer relationship management)-da Data Minning

Data Mining dasturining istiqbolli yo'nalishlaridan biri bu texnologiyani analitik CRM-da qo'llashdir.

CRM (mijozlar bilan munosabatlarni boshqarish) - mijozlar bilan munosabatlarni boshqarish. Ushbu texnologiyalar birgalikda ishlatilganda, data mining va mijozlar ma'lumotlaridan "pul ishlash" bilan birlashtiriladi. Marketing va savdo bo'limlari ishining muhim jihati xaridorlarning yaxlit ko'rinishini, ularning xususiyatlari, xususiyatlari va mijozlar bazasi tuzilishi haqidagi ma'lumotlarni shakllantirishdir. CRM mijozlar haqidagi barcha kerakli ma'lumotlarni to'liq ko'rish imkoniyatini beradigan mijozlar profilining deb nomlangan usulidan foydalanadi. Mijozlarni profillashtirish quyidagi tarkibiy qismlarni o'z ichiga oladi: mijozlarni segmentatsiyalash, mijozlarning rentabelligi, mijozlarni ushlab qolish, mijozlarning javoblarini tahlil qilish. Ushbu tarkibiy qismlarning har biri Data Mining yordamida o'rganilishi mumkin va ularning tahlillari birgalikda, natijada profil komponentlari har bir o'ziga xos xususiyatdan olinmaydigan bilimlarni berishi mumkin. Data Mining-dan foydalanish natijasida mijozlarni rentabelligiga qarab segmentlarga ajratish vazifasi hal qilindi. Tahlil xaridorlarning eng ko'p daromad keltiradigan segmentlarini ta'kidlaydi. Segmentatsiya, shuningdek, mijozlarning sodiqligi asosida amalga oshirilishi mumkin. Segmentatsiya natijasida butun mijozlar bazasi umumiy xususiyatlarga ega bo'lgan ma'lum segmentlarga bo'linadi. Ushbu

xususiyatlarga muvofiq, kompaniya xaridorlarning har bir guruhi uchun marketing siyosatini alohida tanlashi mumkin. Bundan tashqari, Data Mining texnologiyasidan mijozlarning ma'lum bir segmentining ma'lum bir turdagi reklama yoki aktsiyalarga bo'lgan munosabatini taxmin qilish uchun foydalanishingiz mumkin - avvalgi davrlarda to'plangan tarixiy ma'lumotlarga asoslanib. Shunday qilib, Data Mining texnologiyasidan foydalangan holda xaridorlarning xulq-atvorini aniqlash orqali siz marketing, sotish va sotish bo'limlari samaradorligini sezilarli darajada oshirishingiz mumkin. CRM va Data Mining texnologiyalarini va ularni biznesda malakali tatbiq etishni birlashtirib, kompaniya raqobatchilardan sezilarli ustunliklarga ega.

AQSh hukumati mamlakatga kelgan barcha chet elliklarni kuzatib boradigan tizim yaratishni rejalashtirmoqda. Ushbu kompleksning vazifasi: chegara terminalidan boshlab biometrik identifikatsiya qilish texnologiyasi va boshqa turli xil ma'lumotlar bazalari asosida chet elliklarning haqiqiy rejalari ilgari e'lon qilinganlarga (shu jumladan, mamlakat bo'ylab harakatlanish, jo'nab ketish sanasi va hk) mos kelishini nazorat qilish. Tizimning dastlabki qiymati 10 milliard dollardan oshadi, majmuani ishlab chiqaruvchisi - Accenture. AQSh Kongressi Bosh ma'muriyatining analitik hisobotiga ko'ra, AQSh hukumat idoralari ma'lumotlar qazib olishga asoslangan (Data Mining) ikki yuzga yaqin loyihalarda ishtirok etib, aholi to'g'risida turli xil ma'lumotlarni to'playdilar. Ushbu loyihalarning yuzdan ortig'i shaxsiy ma'lumotlarni (ism-shariflar, familiyalar, elektron pochta manzillari, ijtimoiy sug'urta raqamlari va haydovchilik guvohnomalari) to'plashga qaratilgan bo'lib, ushbu ma'lumotlarga asoslanib, odamlar mumkin bo'lgan xatti-harakatlari to'g'risida bashorat qilishadi. Ko'rsatilgan hisobotda maxfiy hisobotlar haqida ma'lumot yo'qligi sababli, bunday tizimlarning umumiy soni ancha ko'p deb taxmin qilish kerak.

Kuzatuv tizimlari olib keladigan afzalliklarga qaramay, ushbu bo'lim mutaxassislari, shuningdek mustaqil ekspertlar bunday loyihalar bilan bog'liq bo'lgan katta xavflar to'g'risida ogohlantiradilar. Xavotirlanish sababi bu kabi bazalarni boshqarish va nazorat qilishda yuzaga kelishi mumkin bo'lgan muammolar.

6. Bioinformatika

Data Mining texnologiyasini qo'llashning ilmiy yo'nalishlaridan biri bu bioinformatika bo'lib, uning yo'nalishi genetik ma'lumotni tahlil qilish va tizimlashtirish algoritmlarini ishlab chiqishdir. Olingan algoritmlar turli xil biologik hodisalarni tushuntirish uchun makromolekulalarning tuzilishini, shuningdek ularning funktsiyalarini aniqlash uchun ishlatiladi.

7. Meditsina

Tibbiyotning ko'plab jihatlari bo'yicha konservatizmiga qaramay, Data Mining texnologiyasi so'nggi yillarda inson faoliyati sohasidagi turli tadqiqotlar uchun faol qo'llanilmoqda. An'anaviy ravishda tibbiy tashxis qo'yish uchun ekspert tizimlari qo'llaniladi, ular ramziy qoidalar asosida qurilgan, masalan, bemorning alomatlari va uning kasalligi. Shablonlar bilan Data Mining-dan foydalanib, siz mutaxassis tizim uchun bilimlar bazasini ishlab chiqishingiz mumkin.

8. Farmatsevtika

Farmatsevtika sohasida Data Mining usullari ham keng qo'llaniladi. Bu ayrim dorilarni klinik foydalanish samaradorligini o'rganish, bemorlarning aniq guruhlari uchun samarali bo'ladigan dorilar guruhlarini aniqlash vazifalari. Dori-darmonlarni bozorga olib chiqish vazifalari bu erda ham dolzarbdir.

Molekulyar genetika va gen injeneriyasi. Molekulyar genetika va genetik muhandislikda Data Mining ning alohida sohasi ajratib olinadi, bu Microarray Data Analysis (MDA) deb nomlanadi. Microarray Data Analysis-dan foydalanish bo'yicha batafsil ma'lumotni bu erda topishingiz mumkin

Ushbu yo'nalishning ba'zi ilovalari:

- erta va aniqroq tashxis qo'yish;
- terapiya uchun yangi molekulyar maqsadlar;
- takomillashtirilgan va individual ravishda davolash usullari;
- fundamental biologik kashfiyotlar.

Data Mining-dan foydalanish misollari - ba'zi bir jiddiy kasalliklarning molekulyar diagnostikasi; genetik kod kasallik ehtimolini oldindan bashorat qilishi mumkinligini kashf qilish; ba'zi yangi dorilar va dorilarning kashf etilishi.

Data Mining tomonidan "Molekulyar genetika va genetik muhandislik" sohalarida ishlatiladigan asosiy tushunchalar markerlar, ya'ni. tirik organizmning turli xususiyatlarini boshqaruvchi genetik kodlar.

Ko'rib chiqilayotgan sohalarda Data Mining yordamida loyihalarni moliyalashtirish uchun muhim moliyaviy resurslar ajratilgan.

Kimyo. Data Mining texnologiyasi organik va noorganik kimyo tadqiqotlarida faol qo'llaniladi. Ushbu sohada Data Mining dasturining mumkin bo'lgan dasturlaridan biri bu minglab elementlarni o'z ichiga olishi mumkin bo'lgan birikmalarning har qanday o'ziga xos konstruktiv xususiyatlarini aniqlashdir. Keyinchalik, biz konchilik yoki "Minning" kontseptsiyasiga asoslangan texnologiyalarni ko'rib chiqamiz.

Web Mining. Web Mining "Internetdagi ma'lumotlarni to`plash" deb tarjima qilinishi mumkin. Web Intelligence yoki Web Intelligent elektron biznesni jadal rivojlantirishda "yangi bob ochishga" tayyor. Har bir tashrif buyuruvchining xatti-harakatlarini kuzatish orqali ularning qiziqishlari va afzalliklarini aniqlash qobiliyati elektron tijorat bozorida jiddiy va tanqidiy raqobat ustunligi hisoblanadi.

Web Mining tizimlari ko'plab savollarga javob berishi mumkin, masalan, tashrif buyuruvchilarning qaysi biri veb-do'konning potentsial xaridoridir, qaysi veb-do'kon mijozlari guruhi ko'proq daromad keltiradi, ma'lum bir mehmon yoki tashrif buyuruvchilar guruhining manfaatlari nimada. Web Minning texnologiyasi sayt ma'lumotlari asosida yangi, ilgari noma'lum bo'lgan bilimlarni kashf etishga qodir bo'lgan va keyinchalik amaliyotda ishlatilishi mumkin bo'lgan usullarni qamrab oladi. Boshqacha qilib aytganda, Web Mining texnologiyasi Data Mining texnologiyasidan foydalanib, veb-saytlarda tuzilgan bo'lmagan, heterojen, tarqatilgan va keng ko'lamli ma'lumotlarni tahlil qiladi. Web Mining taksonomiyasiga ko'ra, bu erda ikkita asosiy yo'nalish mavjud: veb-kontentni olish va veb-foydalanishni olish.

Web Content Mining - bu "axborot shovqini" bilan haddan tashqari yuklangan turli xil Internet-manbalardan yuqori sifatli ma'lumotlarni avtomatik

ravishda qidirish va tanlab olishni anglatadi. Shuningdek, u hujjatlarni klasterlash va izohlash uchun turli xil vositalar bilan shug'ullanadi.

Ushbu yo'nalishda, o'z navbatida, ikkita yondashuv ajratiladi: agentlarga asoslangan yondashuv va ma'lumotlar bazasiga asoslangan yondashuv.

Agentga asoslangan yondashuv quyidagi tizimlarni o'z ichiga oladi:

- aqlli qidiruv agentlari (Intelligent Search Agents);
- axborotni filtrlash / tasniflash;
- shaxsiylashtirilgan tarmoq agentlari.

Aqlli qidiruv agenti tizimlarining misollari:

- Harvest (Braun va boshq., 1994),
- FAQ-Finder (Hammond va boshq., 1995),
- Information Manifold (Kirk va boshq., 1995),
- OCCAM (Kwok and Weld, 1996) va ParaSite (Spertus, 1997),
- ILA (ma'lumotni o'rganish bo'yicha agent) (Perkowitz va Etzioni, 1995),
- ShopBot (Doorenbos va boshq., 1996).

Ma'lumotlar bazalariga yondashuv tizimlarni o'z ichiga oladi:

- ko'p darajali ma'lumotlar bazalari;
- veb-so'rov tizimlari;

Veb-so'rov tizimlarining namunalari:

- W3QL (Konopnicki va Shmueli, 1995),
- WebLog (Lakshmanan va boshq., 1996),
- Lorel (Quass va boshq., 1995),
- UnQL (Buneman va boshq., 1995 va 1996),
- TSIMMIS (Chavathe va boshq., 1994).

Veb-foydalanishni tanlab olishning ikkinchi yo'nalishi veb-sayt foydalanuvchisi yoki ular guruhi harakatlaridagi shakllarni aniqlashni o'z ichiga oladi.

Quyidagi ma'lumotlar tahlil qilinadi:

- foydalanuvchi qaysi sahifalarni ko'rganligi;
- sahifalarni ko'rish ketma-ketligi qanday?

Shuningdek, veb-saytni ko'rish tarixiga asosan foydalanuvchilarning qaysi guruhlarini umumiy sondan ajratish mumkinligi tahlil qilinadi.

Web Usage Mining quyidagi tarkibiy qismlarni o'z ichiga oladi:

- dastlabki qayta berish;
- operatsion identifikatsiya qilish;
- shablonlarni aniqlash vositalari;
- shablonni tahlil qilish vositalari.

Web Miningdan foydalanganda, ishlab chiquvchilar ikkita turdagi vazifalarga duch kelishadi. Birinchisi ma'lumotlar yig'ish bilan bog'liq bo'lsa, ikkinchisi shaxsiylashtirish usullaridan foydalanishga tegishli. Muayyan mijoz haqida ma'lum miqdordagi shaxsiylashtirilgan retrospektiv ma'lumotlarni to'plash natijasida tizim u haqida ma'lum bilimlarni to'playdi va unga, masalan, ma'lum tovar yoki xizmatlarning to'plamlarini tavsiya qilishi mumkin. Sayt barcha tashrif buyuruvchilar haqidagi ma'lumotlarga asoslanib, veb-tizim tashrif buyuruvchilarning aniq guruhlarini aniqlashi, shuningdek ularga mahsulotlarni tavsiya qilishi yoki pochta ro'yxatlaridagi mahsulotlarini taklif qilishi mumkin.

Shunga ko'ra, Web Mining vazifalarini quyidagi toifalarga bo'lish mumkin:

- Veb-konchilik uchun ma'lumotlarni oldindan qayta ishlash.
- Birlashma qoidalari, vaqt ketma-ketliklari, tasniflash va klasterlash yordamida naqshni kashf qilish va bilimlarni kashf etish;
- Olingan bilimlarni tahlil qilish.

Nazorat savollari

1. Web Mining -ning qanday vazifalari mavjud?
2. Aqlli qidiruv tizimlariga misol keltiring.
3. Web Content Mining nima?
4. CRM nima?
5. Data Mining –ning sohalarda qo`llanishiga misollar keltiring.

8-MAVZU

MA`LUMOTLAR TAHLIL ASOSLARI

Reja:

1. **MS Excel da ma`lumotlar tahlili.**
2. **Korellatsion tahlil.**
3. **Regression tahlil.**

Mashg`ulot maqsadi: *Ma'ruza ma'lumotlarni tahlil qilish asoslariga bag'ishlangan, tavsiflovchi statistikaning asosiy xususiyatlarini ko'rib chiqilgan, korrelyatsiya va regressiya tahlilining mohiyatini qisqacha bayon qilingan. Microsoft Excel-da muammolarni echishga misollar keltirilgan.*

Tayanch iboralar: *tavsiflovchi statistika, statistik tahlil, taqdimot, excel, statistik funktsiyalar, tahlillar to'plami, menyu, buyruq, element, statistika, markaziy tendentsiya, o'rtacha, standart xato, o'rtacha, o'rtacha og'ish, dispersiya, ortiqchais, interval, minimal, maksimal, o'rtacha , ishonch oralig'i, diapazon, skewness, haddan tashqari, ustunroq, tahlil, korrelyatsiya koeffitsienti, munosabatlar, juftlik korrelyatsiyasi, regressiya funktsiyasi, qoldiq, munosabatlar darajasi, regressiya koeffitsienti, R-kvadrat, ko'plik R, ma'lumotlar tahlili.*

1. MS Excel da ma`lumotlar tahlili

Ushbu ma'ruzada biz statistik ma'lumotlarni tahlil qilishning ba'zi jihatlarini, xususan tavsiflovchi statistika, korrelyatsiya va regressiya tahlillarini ko'rib chiqamiz. Statistik tahlil juda ko'p turli xil usullarni o'z ichiga oladi, hattoki bitta ma'ruza hajmi juda kichik bo'lgan yuzaki tanishish uchun ham. Ushbu ma'ruzaning maqsadi korrelyatsiya, regressiya tushunchalari to'g'risida eng umumiy tushunchalarni berish, shuningdek tavsiflovchi statistika bilan tanishishdir. Ma'ruzada muhokama qilingan misollar ataylab soddalashtirilgan.

Umumiy maqsadlarga mo'ljallangan paketlar yoki asboblar to'plamlari deb ham ataladigan ko'plab statistik usullarni amalga oshiradigan turli xil amaliy dasturlar to'plami mavjud. Bunday to'plamlar haqida kursning so'nggi qismida

batafsil gaplashamiz. Microsoft Excel shuningdek matematik statistika usullarining keng arsenalini tatbiq etadi, ushbu ma'ruza misollarini amalga oshirish ushbu dasturiy ta'minotda aniq namoyish etildi.

Shuni ta'kidlash kerakki, statistik dasturlardan tashqari, statistik dasturlardan ham foydalanish qiyin - bu uchun foydalanuvchiga maxsus bilim kerak.

Microsoft Excel-da ma'lumotlarni tahlil qilish. Microsoft Excel ko'plab statistik funktsiyalarga ega. Ba'zilari ichki o'rnatilgan, ba'zilari tahlil paketini o'rnatgandan so'ng mavjud. Ushbu ma'ruzada biz ushbu dasturiy ta'minotdan foydalanamiz.

Tahlillar to'plamiga murojaat qilish. Ma'lumotlarni tahlil qilish to'plamiga kiritilgan vositalar Tools menyusining Data Analysis buyrug'i orqali mavjud. Agar menyuda ushbu buyruq bo'lmasa, Asboblar / Qo'shimchalar menyusida "Tahlil to'plami" bandini faollashtirish kerak.

Keyinchalik, tahlil paketiga kiritilgan ba'zi vositalarni ko'rib chiqamiz.

Ta'riflovchi statistika. Ta'riflovchi statistika - bu raqamli ma'lumotlarning massasini tushunish va muhokama qilish oson bo'lgan shaklga aylantirish uchun ishlatiladigan miqdoriy ma'lumotlarni to'plash va umumlashtirish texnikasi. Ta'riflovchi statistikaning maqsadi kuzatishlar va tajribalardan olingan dastlabki natijalarni umumlashtirishdir.

8.1.1-jadvalda berilgan A ma'lumotlar to'plami berilsin.

x	y
3	9
2	7
4	12
5	15
6	17
7	19
8	21
9	23,4
10	25,6
11	27,8

Jadval 8.1.1. A Ma'lumotlar to'plami

Asboblar menyusidan tahlillar to'plamini tanlash va tavsiflovchi statistikani tahlil qilish vositasini tanlash markaziy tendentsiya va kiritilgan ma'lumotlarning o'zgaruvchanligi yoki o'zgarishi to'g'risida ma'lumotlarni o'z ichiga olgan bir o'lchovli statistik hisobotni ishlab chiqaradi. Ta'riflovchi statistika quyidagi xususiyatlarni o'z ichiga oladi: o'rtacha; standart xato; o'rtacha; moda; standart og'ish; namunaviy dispersiya; ortiqcha; assimetriya; oraliq; eng kam; maksimal; miqdor; Xol.Ma'lumotlar to'plami A ning ikkita o'zgaruvchisi uchun tavsiflovchi statistika hisoboti 8.1.2-jadvalda keltirilgan.

	x	y
O'rtacha	6,5	17,68
Standart xato	0,957427108	2,210922382
Mediana	6,5	18
Standart og'ish	3,027650354	6,991550456
Namuna dispersiyasi	9,166666667	48,88177778
Ortiqcha	-1,2	-1,106006058
Asimmetriya	0	-0,128299221
Interval	9	20,8
Eng kam	2	7
Maksimal	11	27,8
Miqdor	65	176,8
Xol	10	10
Eng zo'r (1)	11	27,8
Eng kichik (1)	2	7
Ishonchlilik darajasi (95,0%)	2,16585224	5,001457714

Jadval 8.1.2. Ma'lumotlar to'plami A uchun tavsiflovchi statistika

Ta'riflovchi statistika qanday xususiyatlarga ega ekanligini ko'rib chiqamiz.

Markaziy tendentsiya. Markaziy tendentsiyani o'lchash ma'lumotlar bazasidagi xarakteristikaning barcha qiymatlarini eng yaxshi tavsiflaydigan raqamni tanlashdir. Ushbu raqamning afzalliklari ham, kamchiliklari ham bor. Biz ushbu o'lchovning ikkita xususiyatini ko'rib chiqamiz, ya'ni o'rtacha va o'rtacha, biz keyingi ma'ruzalarda foydalanamiz.

O'rtaning asosiy maqsadi - keyingi tahlil qilish, taqqoslash va taqqoslash uchun ma'lumotlar to'plamini taqdim etish.

O'rtacha osonlik bilan hisoblab chiqiladi va undan keyingi tahlil uchun foydalanish mumkin. Uni intervalli shkalada o'lchangan ma'lumotlar va tartibli shkala bo'yicha o'lchangan ba'zi ma'lumotlar uchun hisoblash mumkin. O'rtacha ma'lumotlar to'plamining o'rtacha arifmetikasi sifatida hisoblanadi: namunadagi barcha qiymatlarning yig'indisi tanlangan hajmga bo'linadi. Ma'lumotlarni shu tarzda "siqish" orqali biz ko'p ma'lumotni yo'qotamiz. O'rtacha juda ma'lumotlidir va o'rganilayotgan barcha ma'lumotlar to'plami to'g'risida xulosa chiqarishga imkon beradi. O'rtacha yordamida biz bir nechta ma'lumotlar to'plamlarini yoki ularning qismlarini taqqoslashimiz mumkin. Ma'lumotlarni tahlil qilishda o'rtacha ko'rsatkichdan ortiqcha foydalanmaslik kerak, uning xususiyatlari va cheklovlarini hisobga olish kerak. "Kasalxonada o'rtacha harorat" yoki "uyning o'rtacha balandligi" xususiyatlari ma'lum bo'lib, ba'zi holatlar uchun markaziy tendentsiyaning ushbu o'lchovidan noto'g'ri foydalanishni ko'rsatmoqda.

O'rtacha xususiyatlar

- O'rtachani hisoblashda yo'qolgan ma'lumotlar qiymatlariga yo'l qo'yilmaydi.
- O'rtacha qiymat faqat raqamli ma'lumotlar va ikkilamchi tarozilar uchun hisoblanishi mumkin.
- Ma'lumotlar to'plami bo'yicha bitta va bitta o'rtacha qiymatni hisoblash mumkin.

O'zgaruvchining o'rtacha qiymatining informatsion qiymati, agar uning ishonch oralig'i ma'lum bo'lsa, yuqori bo'ladi. O'rtacha ishonch oralig'i - bu "haqiqiy" populyatsiya o'rtacha qiymati berilgan ishonch darajasi bilan topilgan taxminiy qiymatlar oralig'i. Ishonch oraliqlarini hisoblash kuzatilgan qiymatlarning normal ekanligi haqidagi taxminlarga asoslanadi. Ishonch oralig'ining kengligi tanlov hajmi va ma'lumotlarning tarqalishiga bog'liq. Tanlov hajmi oshgani sayin o'rtacha bahoning aniqligi oshadi. Tanlangan qiymatlarning tarqalishining ortishi bilan o'rtacha ishonchliligi pasayadi. Agar namunaning kattaligi etarlicha katta bo'lsa, namunaning normal bo'lishidan qat'iy nazar, o'rtacha sifati oshadi.

Mediana - bu kuzatuvlar soniga ko'ra ikkita teng qismga bo'linadigan namunaning aniq o'rta nuqtasi.

Mediani topish uchun zaruriy shart bu namunani buyurtma qilishdir.

Shunday qilib, toq sonli kuzatuvlar uchun median (n + 1) / 2 raqami bilan kuzatuv bo'lib, bu erda n - namunadagi kuzatuvlar soni.

Juft sonli kuzatuvlar uchun median n / 2 va (n + 2) / 2 kuzatuvlarining o'rtacha qiymati hisoblanadi.

Medianing ba'zi xususiyatlari. Ma'lumotlar to'plami bo'yicha bitta o'rtacha qiymatni hisoblash mumkin. Medianani to'liq bo'lmagan ma'lumotlar to'plami uchun ishning raqamlarini tartibda, ishlarning umumiy sonini va ma'lumotlar to'plamining o'rtasida bir nechta qiymatlarni bilish orqali hisoblash mumkin.

Ma'lumotlarning o'zgaruvchanligi xususiyatlari. Namunaning eng oddiy xususiyatlari maksimal va minimaldir.

- **Minimal** - namunadagi eng kichik qiymat.
- **Maksimal** - namunadagi eng katta qiymat.
- **Span** - bu namunadagi eng katta va eng kichik qiymatlar orasidagi farq.
- **Dispersiya** - bu qiymatlarning o'rtacha qiymatidan chetga chiqish kvadratlarining o'rtacha arifmetik qiymati.
- **Standart og'ish**, namunaviy dispersiyaning kvadrat ildizi, ma'lumotlar nuqtalarining o'rtacha qiymatiga nisbatan qanchalik keng tarqalishini o'lchaydi.

Ortiqcha taqsimotning "tepalikning aniqligini" ko'rsatadi, taqsimotning normal taqsimotga nisbatan nisbiy aniqligini yoki silliqligini tavsiflaydi. Ijobiy ortiqcha nisbatan tikanli tarqalishini ko'rsatadi (tepalik ko'rsatiladi). Salbiy ortiqcha nisbatan tekislangan taqsimotni bildiradi (cho'qqisi yumaloq). Agar ortiqcha noldan sezilarli darajada farq qilsa, u holda taqsimot odatdagidan ancha yumaloq cho'qqiga ega, yoki aksincha, keskin tepalikka ega (ehtimol, bir nechta tepaliklar mavjud). Oddiy taqsimotning ortiqchai nolga teng.

Asimmetriya yoki assimetriya taqsimotning nosimmetrikdan og'ishini bildiradi. Agar qiyshiqlik noldan sezilarli darajada farq qilsa, u holda taqsimot assimetrik, normal taqsimot mutlaqo nosimmetrikdir. Agar taqsimot o'ng uzun

quyruqga ega bo'lsa, skewness ijobiy bo'ladi; agar uzun chap quyruq salbiy bo'lsa. Haddan tashqari ko'rsatkichlar - bu ma'lumotlar asosiy qismidan keskin farq qiladigan ma'lumotlar. Haddan tashqari ko'rsatkichlarni aniqlashda tadqiqotchi dilemma bilan duch keladi: tashqi kuzatuvlarni qoldiring yoki ularni rad eting. Ikkinchi variant jiddiy tortishuvlarni va tavsifni talab qiladi. Ma'lumotlarni haddan tashqari ko'rsatkichlar bilan va bo'lmagan holda tahlil qilish va natijalarni taqqoslash foydalidir. Shuni esda tutish kerakki, odatda statistik tahlilning klassik (barqaror) bo'lmagan usullaridan foydalanganda ma'lumotlar bazasida ortiqcha ko'rsatkichlar mavjudligi noto'g'ri natijalarga olib keladi. Ma'lumotlar to'plami nisbatan kichik bo'lsa, ortiqcha hisoblangan ma'lumotlar bundan mustasno, tahlil natijalariga sezilarli ta'sir ko'rsatishi mumkin.

Ma'lumotlar bazasida haddan tashqari ko'rsatkichlarning mavjudligi muntazam xatolar, kirish xatolari, ma'lumotlar yig'ish xatolari va boshqalar bilan bog'liq bo'lgan "bir tomonlama" deb ataladigan qiymatlarning paydo bo'lishi bilan bog'liq bo'lishi mumkin. Ba'zan, ortiqcha ma'lumotlar to'plamidagi eng kichik va eng katta qiymatlarga murojaat qilishlari mumkin.

2. Korrelyatsion tahlil

Korrelyatsion tahlil o'lchovsiz shaklda taqdim etilgan ikkita ma'lumotlar to'plamlari o'rtasidagi munosabatni aniqlash uchun ishlatiladi. Korrelyatsion tahlil ma'lumotlar to'plamlari hajmiga bog'liqligini aniqlashga imkon beradi. Har doim lotincha r harfi bilan belgilanadigan korrelyatsiya koeffitsienti ikkita xususiyat o'rtasida bog'liqlik mavjudligini aniqlash uchun ishlatiladi.

Xususiyatlar o'rtasidagi munosabatlar (Chaddok shkalasi bo'yicha) kuchli, o'rta va zaif bo'lishi mumkin. Ulanishning zichligi korrelyatsiya koeffitsienti qiymati bilan belgilanadi, u qiymatlarni o'z ichiga olgan holda -1 dan +1 gacha olishi mumkin. Aloqa zichligini baholash mezonlari shakl. 8.2.1.

Величина коэффициента корреляции	0.1 - 0.3	0.3 - 0.5	0.5 – 0.7	0.7 – 0.9	0.9 – 1.0
Характеристика силы связи	слабая	умеренная	заметная	высокая	весьма высокая
		средняя		сильная	

Shakl: 8.2.1. Aloqa zichligini baholashning miqdoriy mezonlari

Pirsonning korrelyatsiya koeffitsienti

Pirsonning o'zaro bog'liqlik koeffitsienti r, ya'ni o'lchovsiz indeks bo'lib, shu jumladan -1.0 dan 1.0 gacha, shu jumladan, ikkita ma'lumotlar to'plamlari orasidagi chiziqlilik darajasini aks ettiradi.

Ikki belgi orasidagi bog'liqlikning zichligi ko'rsatkichi chiziqli korrelyatsiya koeffitsienti formulasi bilan aniqlanadi:

$$r_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2] \cdot [n\sum y^2 - (\sum y)^2]}}$$

bu erda **x** - omil atributining qiymati;

y - samarali xususiyatning qiymati;

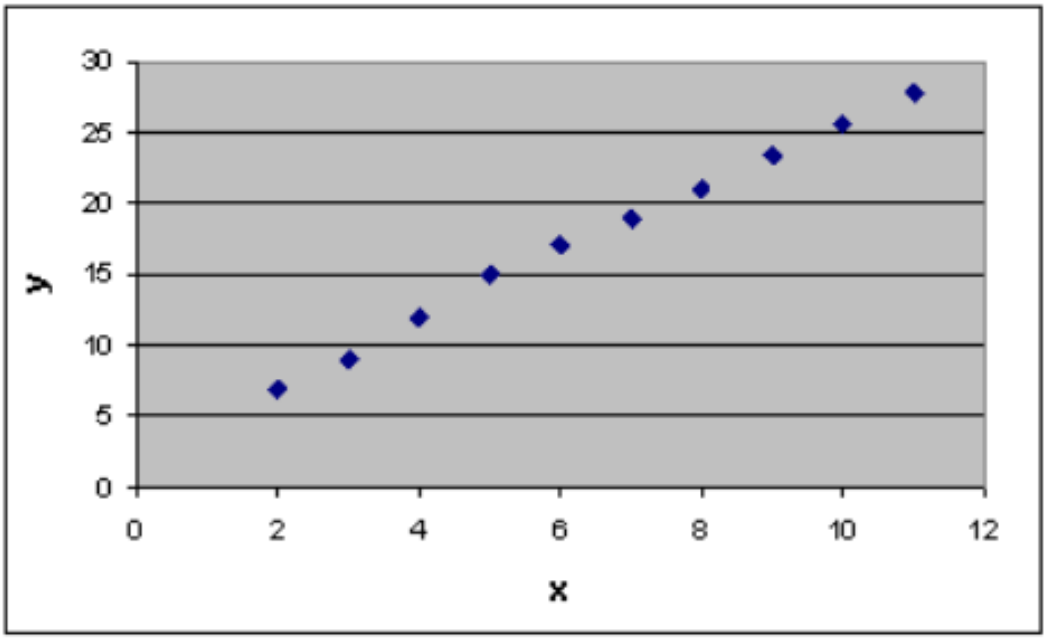
n - ma'lumotlar juftlarining soni.

- bitta ma'lumotlar to'plamidan katta qiymatlar boshqa to'plamning katta qiymatlari bilan bog'liq (ijobiy korrelyatsiya) - to'g'ridan-to'g'ri chiziqli munosabatlarning mavjudligi;
- bir to'plamning kichik qiymatlari boshqasining katta qiymatlari bilan bog'liq (salbiy korrelyatsiya) - manfiy chiziqli munosabat mavjudligi;
- ikki diapazon ma'lumotlari hech qanday bog'liq emas (nol korrelyatsiya) - chiziqli bog'liqlik yo'q.

Misol sifatida A ma'lumotlar to'plamini (8.2.1-jadval) oling. X va y belgilari o'rtasida chiziqli bog'liqlik mavjudligini aniqlash kerak.

Ikki o'zgaruvchining o'zaro bog'liqligini grafik tasvirlash uchun x va y o'zgaruvchilarga mos keladigan o'qlari bo'lgan koordinata tizimi qo'llaniladi. Chiqib ketish uchastkasi deb nomlangan chizilgan uchastka shakl. 8.2.2 Ushbu

diagrammada x ning past qiymatlari y ning past qiymatlariga, x ning yuqori qiymatlari y ning yuqori qiymatlariga mos kelishini ko'rsatadi. Ushbu misol aniq havola mavjudligini ko'rsatadi.



Shakl: 8.2.2. tarqalish diagrammasi

Shu tarzda biz x va y o'zgaruvchilar o'rtasida o'zaro bog'liqlikni o'rnatamiz. MS Excel PEARSON funktsiyasi (massiv1; massiv2) yordamida ikkita massiv (x va y) orasidagi Pearson korrelyatsiya koeffitsientini hisoblaymiz. Natijada biz korrelyatsiya koeffitsientining 0,998364 ga teng qiymatini olamiz, ya'ni. x va y o'zgaruvchilar o'rtasidagi bog'liqlik juda yuqori. MS Excel tahlil to'plami va Korrelyatsion tahlil vositasi yordamida biz korrelyatsiya matritsasini tuzishimiz mumkin. O'zgaruvchilar o'rtasidagi har qanday bog'liqlik ikkita muhim xususiyatga ega: kattalik va ishonchlilik. Ikki o'zgaruvchining o'zaro aloqasi qanchalik kuchli bo'lsa, aloqaning qiymati shunchalik katta bo'ladi va bitta o'zgaruvchining qiymatini boshqa o'zgaruvchining qiymatidan taxmin qilish osonroq bo'ladi. Aloqaning kattaligini ishonchliligidan ko'ra osonroq o'lchash mumkin. Qaramlikning ishonchliligi uning kattaligidan kam emas. Ushbu xususiyat o'rganilayotgan namunaning vakolatliligi bilan bog'liq. Qarama-qarshilikning ishonchliligi ushbu bog'liqlikni boshqa ma'lumotlarda yana topish ehtimolini tavsiflaydi.

O'zgaruvchilarga bog'liqlik qiymati o'sib borishi bilan uning ishonchliligi odatda ortadi.

3. Regression tahlil

Regressiya tahlilining asosiy xususiyati: uning yordami bilan o'rganilayotgan o'zgaruvchilar o'rtasidagi munosabatlar shakli va tabiati to'g'risida aniq ma'lumot olishingiz mumkin.

Regressiyani tahlil qilish bosqichlarining ketma-ketligi

Regressiya tahlilining bosqichlarini qisqacha ko'rib chiqamiz.

1. Muammoni shakllantirish. Ushbu bosqichda o'rganilayotgan hodisalarning bog'liqligi to'g'risida dastlabki taxminlar shakllanadi.
2. Bog'liq va mustaqil (tushuntiruvchi) o'zgaruvchilarni aniqlash.
3. Statistik ma'lumotlar to'plami. Regressiya modeliga kiritilgan o'zgaruvchilarning har biri uchun ma'lumotlar to'planishi kerak.
4. Aloqa shakli (oddiy yoki ko'p, chiziqli yoki chiziqli bo'lmagan) haqida gipotezani shakllantirish.
5. Regressiya funktsiyasini aniqlash (regressiya tenglamasi parametrlarining son qiymatlarini hisoblashdan iborat)
6. Regressiya tahlilining aniqligini baholash.
7. Olingan natijalarni talqini. Regressiya tahlilining olingan natijalari dastlabki gipotezalar bilan taqqoslanadi. Olingan natijalarning to'g'riligi va ehtimoli baholanadi.
8. Qaram o'zgaruvchining noma'lum qiymatlarini bashorat qilish.

Regressiya tahlili yordamida bashorat qilish va tasniflash masalasini hal qilish mumkin. Bashorat qilingan qiymatlar parametr regressiya tenglamasiga tushuntiruvchi o'zgaruvchan qiymatlarni almashtirish orqali hisoblanadi. Tasniflash masalasi quyidagi tarzda echiladi: regressiya chizig'i barcha ob’yektlar to'plamini ikkita sinfga ajratadi va funktsiya qiymati noldan katta bo'lgan to'plamning bu qismi bitta sinfga tegishli bo'lib, u joylashgan joy. noldan kam boshqa sinfga tegishli.

Regressiyani tahlil qilish vazifalari. Regressiya tahlilining asosiy vazifalarini ko'rib chiqamiz: qaramlik shaklini o'rnatish, regressiya funktsiyasini aniqlash, qaram o'zgaruvchining noma'lum qiymatlarini baholash.

Bog`liklik shaklini belgilash. O'zgaruvchilar o'rtasidagi bog'liqlikning tabiati va shakli quyidagi regressiya turlarini hosil qilishi mumkin:

- ijobiy chiziqli regressiya (funktsiyaning bir xil o'sishida ifodalangan);
- ijobiy bir xilda ortib borayotgan regressiya;
- ijobiy, barqaror o'sib borayotgan regressiya;
- salbiy chiziqli regressiya (funktsiyaning bir tekis tushishi bilan ifodalangan);
- salbiy bir xil kamayib boruvchi regressiya;
- salbiy, teng darajada sekinlashgan regressiya.

Biroq, ta'riflangan navlar odatda sof shaklda emas, balki bir-biri bilan birgalikda topiladi. Bunday holda, kishi regressiyaning birlashgan shakllari haqida gapiradi.

Regressiya funktsiyasini aniqlash.

Regressiya tenglamasi. Ushbu tenglama Y o'zgaruvchisini a o'zgaruvchisi va chiziqning (yoki qiyalikning) qiyaligi b ning X o'zgaruvchisidan qiymatiga nisbatan ifodalaydi.A doimiyligi kesma deb ham ataladi va qiyalik regressiya koeffitsienti yoki B koeffitsientidir. Ko'pgina hollarda (har doim ham bo'lmasa), regressiya chizig'iga nisbatan kuzatuvlarning ma'lum bir tarqalishi mavjud.

Qoldiq - bitta nuqtaning (kuzatuvning) regressiya chizig'idan (taxmin qilingan qiymat) chetga chiqishi.

MS Excel-da regressiya tahlili masalasini hal qilish uchun Asboblar menyusidan "Tahlillar to'plami" va "Regressiya" tahlil vositasini tanlang. Biz X va Y kirish diapazonlarini o'rnatdik. Y kirish oralig'i tahlil qilinadigan bog'liq ma'lumotlar doirasidir, u bitta ustunni o'z ichiga olishi kerak. Kirish oralig'i - bu tahlil qilinishi kerak bo'lgan mustaqil ma'lumotlar doirasi. Kirish diapazonlari soni 16 dan oshmasligi kerak.

Chiqish diapazonida protsedura chiqarilganda biz 8.3.1.a - 8.3.1.b jadvalda keltirilgan hisobotni olamiz.

Natija:

Regressiya statistikasi	
Ko`p sonli R	0,998364
R-kvadrat	0,99673
Normallashtirilgan R-kvadrat	0,996321
Standart xatolik	0,42405
Kuzatuv	10

Jadval 8.3.1.a. Regressiya statistikasi

Birinchidan, 8.3.1.a-jadvalda keltirilgan hisob-kitoblarning yuqori qismini - regressiya statistikasini ko'rib chiqamiz.

R-kvadratik qiymat, shuningdek, aniqlik o'lchovi deb ataladi, natijada paydo bo'ladigan regressiya chizig'ining sifatini tavsiflaydi. Ushbu sifat dastlabki ma'lumotlar va regressiya modeli (hisoblangan ma'lumotlar) o'rtasidagi moslik darajasi bilan ifodalanadi. Ishonchlilik o'lchovi har doim [0; 1] oralig'ida bo'ladi.

Ko'pgina hollarda, R-kvadrat qiymati ushbu qiymatlar orasida, haddan tashqari deb nomlanadi, ya'ni. nol va bitta o'rtasida.

Agar R kvadratik qiymati biriga yaqin bo'lsa, demak, tuzilgan model mos keladigan o'zgaruvchilarning deyarli barcha o'zgaruvchanligini tushuntiradi. Aksincha, nolga yaqin bo'lgan R kvadratik qiymati qurilgan modelning sifatsizligini anglatadi.

Bizning misolimizda aniqlik o'lchovi 0.99673 ni tashkil etadi, bu regressiya chizig'ining dastlabki ma'lumotlarga juda mos kelishini ko'rsatadi.

Ko'p sonli R - ko'p korrelyatsiya koeffitsienti R - mustaqil o'zgaruvchilar (X) va bog'liq o'zgaruvchiga (Y) bog'liqlik darajasini ifodalaydi.

Ko'p sonli R aniqlanish koeffitsientining kvadrat ildiziga teng; bu qiymat noldan birgacha bo'lgan qiymatlarni oladi.

Oddiy chiziqli regressiya tahlilida R ko'pligi Pirsonning korrelyatsiya koeffitsientiga teng. Darhaqiqat, bizning holatimizdagi R oldingi misoldan (0.998364) Pirson korrelyatsiya koeffitsientiga teng.

	Koeffisentlar	Standart xatolik	t-statistika
Y-kesishma	2,694545455	0,33176878	8,121757129
O`zgaruvchi X 1	2,305454545	0,04668634	49,38177965
*** Hisob-kitoblarning qisqartirilgan versiyasi ko'rsatilgan**			

Jadval 8.3.1.b. Regressiya koeffitsientlari

Endi 8.3.1.b-jadvalda keltirilgan hisob-kitoblarning o'rta qismini ko'rib chiqamiz. Bu erda regressiya koeffitsienti b (2.305454545) va ordinat bo'ylab siljish berilgan, ya'ni. doimiy a (2.694545455).

Hisob-kitoblarga asoslanib, regressiya tenglamasini quyidagicha yozishimiz mumkin:

Y= x*2,305454545+2,694545455

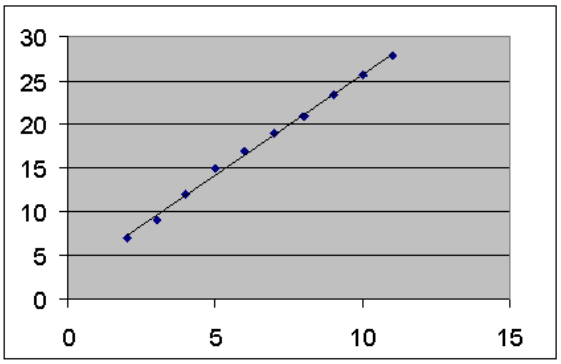
O'zgaruvchilar o'rtasidagi bog'liqlik yo'nalishi regressiya koeffitsientlari (b koeffitsienti) belgilariga (salbiy yoki musbat) asoslangan holda aniqlanadi. Agar regressiya koeffitsientining belgisi ijobiy bo'lsa, bog'liq o'zgaruvchi va mustaqil o'zgaruvchining o'zaro munosabati ijobiy bo'ladi. Bizning holatimizda regressiya koeffitsientining belgisi ijobiy, shuning uchun munosabatlar ham ijobiydir. Agar regressiya koeffitsientining belgisi salbiy bo'lsa, bog'liq o'zgaruvchi va mustaqil o'zgaruvchining o'zaro bog'liqligi salbiy (teskari) bo'ladi. Jadval 8.3.1v. qoldiqlarni olib tashlash natijalari keltirilgan. Ushbu natijalar hisobotda ko'rinishi uchun, "Regression" vositasini ishga tushirganda, "Qoldiqlar" katakchasini faollashtirishingiz kerak.

Qoldiqni chiqarish:

Kuzatish	Bashorat qilingan Y	Qoldiq	Standart qoldiqlar
1	9,610909091	-0,610909091	-1,528044662
2	7,305454545	-0,305454545	-0,764022331
3	11,91636364	0,083636364	0,209196591
4	14,22181818	0,778181818	1,946437843
5	16,52727273	0,472727273	1,182415512
6	18,83272727	0,167272727	0,418393181
7	21,13818182	-0,138181818	-0,34562915
8	23,44363636	-0,043636364	-0,109146047
9	25,74909091	-0,149090909	-0,372915662
10	28,05454545	-0,254545455	-0,636685276

Jadval 8.3.1.v. Qoldiqlar

Hisobotning ushbu qismi yordamida biz har bir nuqtaning chizilgan regressiya chizig'idan chetga chiqishini ko'rishimiz mumkin. Qoldiqning eng katta mutlaq qiymati bizning holatimizda 0,778, eng kichigi 0,043 ga teng. Ushbu ma'lumotlarni yaxshiroq talqin qilish uchun biz dastlabki ma'lumotlarning grafigi va shaklda keltirilgan qurilgan regressiya chizig'idan foydalanamiz. 8.3.1.-shakl. Ko'rib turganingizdek, regressiya chizig'i asl ma'lumotlarning qiymatlariga to'liq mos keladi. Shuni yodda tutish kerakki, ko'rib chiqilayotgan misol juda sodda va yuqori sifatli chiziqli regressiya chizig'ini qurish har doim ham mumkin emas.



Shakl: 8.3.1. Dastlabki ma'lumotlar va regressiya chizig'i

Mustaqil o'zgaruvchining ma'lum qiymatlari asosida qaram o'zgaruvchining kelajakdagi noma'lum qiymatlarini baholash muammosi ko'rib chiqilmagan bo'lib qoldi, ya'ni. prognozlash muammosi.

Regressiya tenglamasiga ega bo'lib, prognozlash muammosi ma'lum x qiymatlari bilan Y = x * 2.305454545 + 2.694545455 tenglamasini echishga kamayadi. Olti qadam oldinda bog'liq bo'lgan o'zgaruvchini bashorat qilish natijalari 8.3.2.-jadvalda keltirilgan.

X	Y(Bashorat qilingan)
11	28,05455
12	30,36
13	32,66545
14	34,97091
15	37,27636
16	39,58182

8.3.2.-jadval. Y o'zgaruvchisini bashorat qilish natijalari

Shunday qilib, Microsoft Excel paketida regressiya tahlilidan foydalanish natijasida biz:

- regressiya tenglamasini tuzildi;
- bog`liqlik shakli va o'zgaruvchilar o'rtasidagi bog'liqlik yo'nalishini o'rnatdik - funktsiyaning bir xil o'sishida ifodalangan ijobiy chiziqli regressiyani aniqladik;
- o'zgaruvchilar o'rtasidagi bog'liqlik yo'nalishini o'rnatdik;
- olingan regressiya chizig'i sifatini baholandi;
- hisoblangan ma'lumotlarning asl to'plam ma'lumotlaridan chetga chiqishini ko'rishga muvaffaq bo'ldik;
- bog`liq o'zgaruvchining kelajakdagi qiymatlarini bashorat qildik.

Agar regressiya funktsiyasi aniqlangan, talqin qilingan va asoslangan bo'lsa va regressiya tahlilining aniqligini baholash talablarga javob bersa, biz tuzilgan model va taxmin qilingan qiymatlar etarli ishonchliligiga ega deb taxmin qilishimiz mumkin.Shu tarzda olingan prognoz qiymatlari kutish mumkin bo'lgan o'rtacha qiymatlardir.

XULOSA. Mashg`ulotimizning ushbu qismida biz tavsiflovchi statistikaning asosiy xususiyatlarini ko'rib chiqdik va ular orasida ma'lumotlarning o'rtacha, o'rtacha, maksimal, minimal va boshqa xususiyatlari kabi tushunchalar ko'rib chiqildi. Emissiya tushunchasi ham qisqacha muhokama qilindi. Ma'ruzada muhokama qilingan xususiyatlar kashfiyot ma'lumotlari tahlili deb ataladi, uning xulosalari umumiy aholiga taalluqli emas, faqat ma'lumotlarning namunalariga tegishli bo'lishi mumkin. Ma'lumotlarni qidiruv tahlili dastlabki xulosalar chiqarish va aholi to'g'risida farazlarni shakllantirish uchun ishlatiladi. Shuningdek, korrelyatsiya va regressiya tahlili asoslari, ularning vazifalari va amaliy foydalanish imkoniyatlari ko'rib chiqildi.

Nazorat savollari

1. Regressiya nima? MIsol keltiring.
2. Pirson korrelyatsiyasini tushuntirib bering.
3. Regressiyani tahlil qilishning qanday bosqichlarini bilasiz?
4. Korrelyatsiya nima?

9-MAVZU

KLASSIFIKATSIYALASH VA PROGNOZLASH METODLARI. YECHIMLAR DARAXTI

Reja:

1. **Klassifikatsiyalash va prognozlash metodlari.**
2. **Yechimlar daraxti**

Mashg`ulot maqsadi: *Yechim daraxtlari usuli tasvirlangan. Yechim daraxtining elementlari, uni qurish jarayoni ko'rib chiqiladi. Klassifikatsiya masalasini yechadigan daraxtlarga misollar keltirilgan. CART va C4.5 yechim daraxtlarini qurish algoritmlari berilgan.*

Tayanch iboralar: *yechim daraxti, Ma'lumotlarni tanlab olish, qoida, o'zgaruvchan, daraxt, induksiya, o'yin echimi, tekshiruv tuguni, filial, ichki tugun, yakuniy tugun, qaror tuguni, ta'rif, tekshiruv tuguni, barg, tepa, ikkilik tasnif, filial, ma'lumotlar bazasi, kredit , yo'l, ma'lumotlar bazalari, bo'linish atributi, atribut, cheklangan, bo'linish predikati, yozuv, ma'lumot, bo'linish mezoni, mezon, ob'yekt, algoritm, algoritm kiritish, statistik usullar, o'lchovlilik, foydalanuvchi, nazoratsiz o'rganish, yaratish, daraxt, qisqartirish, kesish, daromad , arava, tasnif, regressiya, LEO, ikkilik yechimlar daraxti, minimallashtirish, qoidalar.*

1. Klassifikatsiyalash va prognozlash metodlari.

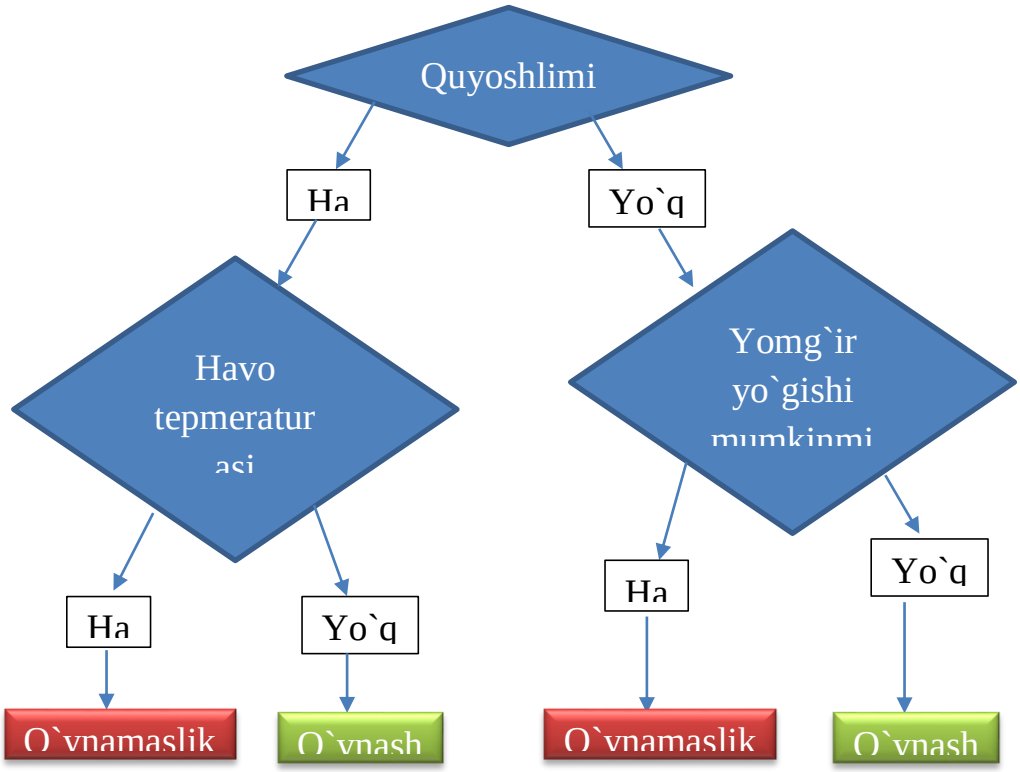
Yechim daraxtlari tasniflash va bashorat qilish muammolarini hal qilishning eng mashhur usullaridan biridir. Ba'zida bu Data Mining usuli, shuningdek, qaror qoidalari daraxtlari, tasnif daraxtlari va regressiya daraxtlari deb nomlanadi. Familiyadan ko'rinib turibdiki, bu usul tasniflash va bashorat qilish muammolarini hal qilishda ishlatiladi. Agar bog`liqlik bo'lsa, ya'ni, maqsadli o'zgaruvchi diskret qiymatlarni oladi, tasniflash muammosi yechim daraxti usuli yordamida hal qilinadi. Agar bog`liqlik o'zgaruvchi uzluksiz qiymatlarni qabul qilsa, u holda yechimlar daraxti ushbu o'zgaruvchining mustaqil o'zgaruvchilarga bog'liqligini o'rnatadi, ya'ni, raqamli bashorat qilish masalasini hal qiladi. Yechim daraxtlari birinchi bo'lib

1950-yillarning oxirlarida Xovelend va Xant tomonidan taklif qilingan. Yechim daraxtlarining mohiyatini bayon etgan Xant va boshqalarning eng qadimgi va taniqli asari "Induksiyadagi tajribalar" 1966 yilda nashr etilgan.

Eng sodda shaklda yechim daraxti - bu ierarxik, ketma-ket tuzilishda qoidalarni ifodalash usuli. Ushbu tuzilmaning asosi qator savollarga "Ha" yoki "Yo'q" javobidir.

Shakl. 9.1.1. da yechim daraxtining namunasini ko'rsatadi, uning vazifasi: "Men golf o'ynashim kerakmi?" Degan savolga javob berishdir. Muammoni hal qilish uchun, ya'ni, golf o'ynash yoki qilmaslik to'g'risida qaror qabul qilish uchun mavjud vaziyatni ma'lum bo'lgan sinflardan biriga bog'lash kerak (bu holda "o'ynash" yoki "o'ynamaslik"). Buning uchun ushbu daraxtning ildizidan boshlab tugunlarida joylashgan bir qator savollarga javob berish kerak.

Daraxtimizning birinchi tuguni "Quyoshli?" tasdiqlash tuguni, ya'ni. holat. Agar savol ijobiy bo'lsa, chap filial deb nomlangan daraxtning chap tomoniga o'tish, agar javob salbiy bo'lsa, daraxtning o'ng tomoniga amalga oshiriladi. Shunday qilib, daraxtning ichki tuguni ma'lum bir holatni tekshirish tugunidir.



Shakl: 9.1.1. Yechim daraxti "Golf o'ynashim kerakmi?"

Keyingi navbatdagi savol keladi va hokazo, yakuniy daraxt tuguniga yetguncha, bu qaror tuguni. Daraxtimiz uchun ikkita tugun mavjud: golfni "o'ynash" va "o'ynamaslik". Daraxtning ildizidan (ba'zan uni ildiz tepasi deb ham atashadi) uning tepasiga o'tish natijasida tasniflash masalasi hal qilinadi, ya'ni, sinflardan biri tanlanadi – golfni "o'ynash" va "o'ynamaslik".Bizning holatimizda yechim daraxtini barpo etishdan maqsad, toifaga bog'liq o'zgaruvchining qiymatini aniqlashdir. Shunday qilib, bizning vazifamiz uchun yechim daraxtining asosiy elementlari: Daraxt ildizi: "Quyoshli?"

Ichki daraxt tuguni yoki sinov tuguni: "Havoning harorati balandmi?", "Yomg'ir yog'yaptimi?" Barg, daraxtning so'nggi nuqtasi, echim tuguni yoki tepalik: O`yna, O`ynamang. Daraxt novdasi (javob holatlari): "Ha", "Yo'q".

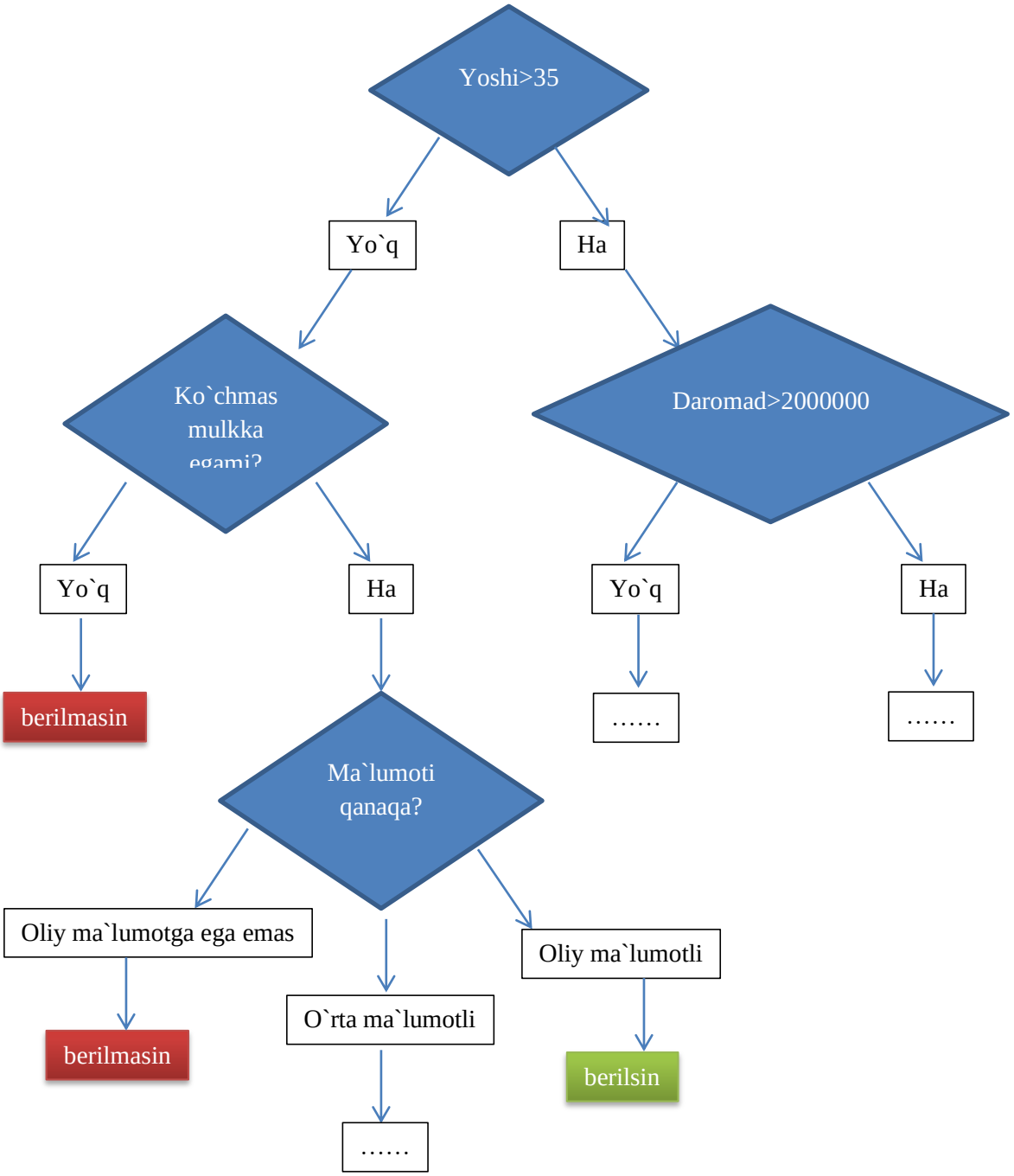
Ko'rib chiqilgan misol ikkilik tasniflash muammosini hal qiladi, ya'ni. ikkilamchi tasniflash modeli yaratildi. Misol, ikkilik daraxtlar deb nomlangan ishlarni namoyish etadi. Ikkilik daraxtlarning tugunlarida dallanish faqat ikki yo'nalishda amalga oshirilishi mumkin, ya'ni. berilgan savolga faqat ikkita javob ("ha" va "yo'q") imkoniyati mavjud.

Ikkilik daraxtlar - bu eng oddiy, alohida qaror qilingan daraxtlar. Boshqa hollarda, ikkitadan ortiq javob bo'lishi mumkin va shunga ko'ra daraxtning ichki tugunidan chiqqan shoxlari.

Keling, yanada murakkab bir misolni ko'rib chiqaylik. Bashoratni amalga oshirish kerak bo'lgan ma'lumotlar bazasida bankning mijozlari to'g'risida quyidagi atributlar mavjud bo'lgan retrospektiv ma'lumotlar mavjud: yoshi, ko'chmas mulk mavjudligi, ma'lumoti, o'rtacha oylik daromadi, mijoz kreditni to'laganmi yoki yo'qmi. vaqt. Vazifa yuqorida sanab o'tilgan ma'lumotlarga asoslanib (oxirgi atributdan tashqari) yangi mijozga qarz berishga arziydimi yoki yo'qligini aniqlashdan iborat. Tasniflash muammosi bo'yicha ma'ruzada ko'rib chiqqanimizdek, bunday muammo ikki bosqichda hal qilinadi: tasniflash modelini yaratish va undan foydalanish.

Modelni qurish bosqichida, aslida, tasnif daraxti quriladi yoki ma'lum qoidalar to'plami yaratiladi. Modeldan foydalanish bosqichida "qarz berishim

kerakmi?" Degan savolga javob berish uchun qurilgan daraxt yoki ma'lum bir mijoz uchun qoidalar to'plami bo'lgan uning ildizidan tepalikka biriga yo'l ishlatiladi.



Shakl: 9.1.2. Yechim daraxti "Qarz beris kerakmi?"

Qoida "agar: keyin:" shaklida taqdim etilgan mantiqiy qurilishdir. Shakl. 9.1.2. "Mijozga qarz berishim kerakmi?" muammosini hal qilish uchun foydalaniladigan tasnif daraxtining namunasi ko'rsatilgan. Bu odatiy tasniflash muammosi va yechim daraxtlari yordamida juda yaxshi echimlar olinadi.

Ko'rib turganimizdek, daraxtning ichki tugunlari (yoshi, ko'chmas mulk, daromad va ma'lumot) yuqorida tavsiflangan ma'lumotlar bazasining atributlari hisoblanadi. Ushbu atributlar bashorat qiluvchi yoki bo'linadigan atributlar deb ataladi. Daraxtning bargli tugunlari yoki barglari, "qarz berish" yoki "qarz bermaslik" kreditiga bog'liq bo'lgan toifali o'zgaruvchining qiymatlari bo'lgan sinf yorliqlari deb ataladi.

Ichki tugundan daraxtning har bir novdasi bo'linish predikati bilan belgilanadi. Ikkinchisi berilgan tugunning faqat bitta bo'linish atributiga murojaat qilishi mumkin. Bo'linish predikatlarining o'ziga xos xususiyati shundaki, har bir yozuv daraxt ildizidan bittagina echim tugunigacha bo'lgan noyob yo'lni ishlatadi. Atributlarni ajratish va predikatlar tugunida bo'linish haqida birlashtirilgan ma'lumot bo'linish mezoni deb ataladi.

Shakl. 9.1.2. da ko'rib chiqilayotgan ma'lumotlar bazasi uchun mumkin bo'lgan yechim daraxtlaridan birini tasvirlaydi. Masalan, "Qanday ta'lim?" Bo'linish mezoni ikkita bo'linishga ega bo'lishi va boshqacha ko'rinishga ega bo'lishi mumkin: "oliy" va "yuqori bo'lmagan" ta'lim. Shunda qaror daraxti boshqacha ko'rinishga ega bo'lar edi.

Shunday qilib, ma'lum bir muammo uchun (shuningdek, boshqa har qanday narsa uchun) har xil taxminiy aniqlik bilan har xil sifatli qaror daraxtlari to'plamini qurish mumkin.

Qurilgan qaror daraxtining sifati bo'linish mezonining to'g'ri tanlanishiga juda bog'liq. Ko'pgina tadqiqotchilar mezonlarni ishlab chiqish va takomillashtirish ustida ishlamoqdalar.

Yechim daraxti usuli ko'pincha "sodda" yondashuv deb nomlanadi. Ammo bir qator afzalliklar tufayli ushbu usul tasniflash muammolarini hal qilishda eng mashhurlaridan biri hisoblanadi.

2. Yechimlar daraxti

Intuitiv qaror daraxtlari. Yechim daraxti shaklida taqdim etilgan tasniflash modeli intuitiv bo'lib, yechilayotgan muammoni tushunishni soddalashtiradi. Yechim daraxtlarini qurish algoritmlari ishining natijasi, masalan, "qora qutilar"

bo'lgan neyron tarmoqlardan farqli o'laroq, foydalanuvchi tomonidan osonlikcha izohlanadi. Yechim daraxtlarining bu xususiyati yangi ob’yektni ma'lum bir sinfga berishda nafaqat muhim, balki tasniflash modelini bir butun sifatida talqin qilishda ham foydalidir. Yechim daraxti nima uchun ma'lum bir ob’yekt ma'lum bir sinfga tegishli ekanligini tushunishga va tushuntirishga imkon beradi.

Yechim daraxtlari tabiiy til ma'lumotlar bazasidan qoidalarni chiqarib olish imkoniyatini beradi. Namunaviy qoida: Agar yosh> 35 va daromad> 2000000 bo'lsa, unda qarz bering. Yechim daraxtlari analitik uchun bilimlarni rasmiylashtirishi qiyin bo'lgan joylarda tasniflash modellarini yaratishga imkon beradi. Yechimlar daraxtini loyihalash algoritmi foydalanuvchidan kirish atributlarini (mustaqil o'zgaruvchilar) tanlashini talab qilmaydi. Algoritmga mavjud bo'lgan barcha atributlarni kiritish mumkin, algoritmning o'zi ular orasida eng muhimini tanlaydi va faqat ular daraxtni qurish uchun ishlatiladi. Masalan, neyron tarmoqlari bilan taqqoslaganda, bu foydalanuvchi ishini ancha osonlashtiradi, chunki neyron tarmoqlarida kirish atributlari sonini tanlash mashg'ulot vaqtiga sezilarli ta'sir qiladi.

Yechim daraxtlari yordamida yaratilgan modellarning aniqligi tasniflash modellarini yaratishning boshqa usullari (statistik usullar, neyron tarmoqlari) bilan taqqoslanadi. Ultra katta ma'lumotlar bazalarida yrchim daraxtlarini qurish uchun ishlatilishi mumkin bo'lgan bir qator o'lchovli algoritmlar ishlab chiqilgan; bu erda ko'lamlilik shuni anglatadiki, misollar yoki ma'lumotlar bazasi yozuvlari ko'payib borishi bilan mashg'ulotga sarflanadigan vaqt, ya'ni. chiziqli o'sadigan qaror daraxtlarini qurish. Bunday algoritmlarga misollar: SLIQ, SPRINT.

Tez o'rganish jarayoni. Yechim daraxtini loyihalash algoritmlaridan foydalangan holda tasniflash modellarini yaratish, masalan, neyron tarmoqlarni o'qitishdan ancha kam vaqt talab etadi. Yechimlar daraxtini loyihalash algoritmlarining aksariyati etishmayotgan qiymatlarni maxsus usulda boshqarish qobiliyatiga ega. Tasniflash muammolarini hal qiladigan ko'plab klassik statistik usullar faqat raqamli ma'lumotlar bilan ishlashi mumkin, qaror daraxtlari esa raqamli va kategorik ma'lumotlar turlari bilan ishlaydi.

Ko'pgina statistik usullar parametrli bo'lib, foydalanuvchi oldindan ma'lum ma'lumotlarga ega bo'lishi kerak, masalan, model turini bilishi, o'zgaruvchilar o'rtasidagi bog'liqlik turi to'g'risida farazga ega bo'lishi va ma'lumotlarning qanday taqsimlanishiga ega ekanligini taxmin qilishi kerak. Yechim daraxtlari, bunday usullardan farqli o'laroq, parametrik bo'lmagan modellarni yaratadilar. Shunday qilib, qaror daraxtlari o'rganilayotgan ma'lumotlar orasidagi bog'liqlik turi to'g'risida apriori ma'lumot bo'lmagan Data Mining kabi muammolarni hal qilishga qodir.

Yechimlar daraxtini loyihalash jarayoni. Eslatib o'tamiz, biz ko'rib chiqayotgan tasniflash muammosi ba'zan induktiv ta'lim deb ataladigan nazorat ostida o'qitish strategiyasiga tegishli. Bunday hollarda, o'quv ma'lumotlar bazasidagi barcha ob’yektlar oldindan belgilangan sinflardan biriga tayinlangan.

Yechim daraxtlarini qurish algoritmlari daraxtni "qurish" yoki "yaratish" (daraxt qurish) va daraxtni "kesish" (daraxtlarni kesish) bosqichlaridan iborat. Daraxt yaratish jarayonida bo'linish va mashg'ulotni to'xtatish mezonini tanlash masalalari hal qilinadi (agar algoritmda nazarda tutilgan bo'lsa). Daraxtlarni qisqartirish bosqichida uning ba'zi shoxlarini kesish masalasi hal qilinadi. Keling, ushbu masalalarni batafsil ko’rib chiqaylik.

Bo’linish mezoni. Daraxt yuqoridan pastgacha yaratilgan, ya'ni. yuqoridan pastga. Jarayon davomida algoritm to'plamni berilgan sinov tuguni bilan bog'liq bo'lgan pastki qismlarga ajratish uchun shunday bo'linish mezonini topishi kerak, ba'zan uni ajratish mezonlari deb ham atashadi. Har bir sinov tuguni ma'lum bir atribut bilan belgilanishi kerak. Atributni tanlash qoidasi bor: u dastlabki ma'lumotlar to'plamini shunday ajratishi kerakki, bu bo'linish natijasida olingan kichik to'plamlarning ob’yektlari bir xil sinf vakillari yoki bunday bo'linishga imkon qadar yaqinroq bo'lsin. So'nggi ibora shuni anglatadiki, har bir sinfdagi "iflosliklar" deb nomlangan boshqa sinflarning ob’yektlari soni minimal darajaga tushishi kerak.

Turli xil bo'linish mezonlari mavjud. Eng mashhurlari entropiya o'lchovi va **Gini indeksi.** Ayrim usullarda entropiya yondashuviga asoslangan va ma'lumot olish o'lchovi yoki entropiya o'lchovi sifatida tanilgan bo'linish atributini tanlash uchun atributlarning axborot yutug'i ishlatiladi.

Breiman va boshqalar tomonidan taklif qilingan yana bir bo'linish mezonlari.CART algoritmida amalga oshiriladi va Gini indeksi deb nomlanadi. Ushbu indeks bilan atribut sinf taqsimotlari orasidagi masofaga qarab tanlanadi.

N sinflari misollarini o'z ichiga olgan T to'plami berilgan, Gini indeksi, ya'ni. gini (T), quyidagi formula bilan aniqlanadi:

$$gini(T) = 1 - \sum_{j=1}^{n} p_j^2$$

bu erda T - joriy tugun, pj - T tugunidagi j sinfining ehtimoli, n - sinflar soni.

Katta daraxt bu "yaroqli" degani emas

Yechim daraxtida qancha ko'p maxsus holatlar tasvirlangan bo'lsa, har bir alohida holatga kamroq ob’yektlar tushadi. Bunday daraxtlar "tarvaqaylab ketgan" yoki "butali" deb nomlanadi, ular asossiz ravishda ko'p sonli tugun va shoxlardan iborat bo'lib, asl to'plam juda oz sonli ob’yektlardan tashkil topgan juda ko'p kichik to'plamlarga bo'linadi. Bunday daraxtlarning "toshib ketishi" natijasida ularning umumlashtirish qobiliyati pasayadi va qurilgan modellar to'g'ri javob bera olmaydi.

Daraxtni qurish jarayonida uning kattaligi haddan tashqari kattalashib ketmasligi uchun "mos o'lchamdagi" daraxtlar deb ataladigan maqbul daraxtlarni yaratishga imkon beradigan maxsus protseduralar qo'llaniladi (Breiman, 1984).

Daraxtning optimal hajmi qanday? Daraxt o'rganilayotgan ma'lumotlar to'plamidagi ma'lumotlarni hisobga oladigan darajada murakkab bo'lishi kerak, ammo shu bilan birga u etarlicha sodda bo'lishi kerak. Boshqacha qilib aytganda, daraxt modelning sifatini yaxshilaydigan ma'lumotlardan foydalanishi va uni yaxshilamaydigan ma'lumotlarga e'tibor bermasligi kerak.

Bu erda ikkita strategiya mavjud. Birinchisi, daraxtni foydalanuvchi tomonidan belgilangan parametrlarga muvofiq ma'lum hajmga etishtirishdir. Ushbu parametrlarni aniqlash tahlilchining tajribasi va sezgisiga, shuningdek, qarorlar daraxtini barpo etuvchi tizimning ba'zi "diagnostik xabarlariga" asoslangan bo'lishi mumkin.

Ikkinchi strategiya Briman, Quiland va boshq tomonidan ishlab chiqilgan daraxtning "mos o'lchamini" aniqlash uchun protseduralar to'plamidan foydalanishdir.1984 yilda. Biroq, mualliflar ta'kidlaganidek, ushbu protseduralar yangi boshlagan foydalanuvchi uchun mavjud deb aytish mumkin emas.

Haddan tashqari katta daraxtlarning paydo bo'lishiga yo'l qo'ymaslik uchun qo'llaniladigan protseduralarga quyidagilar kiradi: daraxtni shoxlarini kesib kesish; o'rganishni to'xtatish qoidalaridan foydalanish.

Shuni ta'kidlash kerakki, daraxtni qurish uchun barcha algoritmlar bir xil sxema bo'yicha ishlamaydi. Ba'zi algoritmlar ketma-ket ikkita bosqichni o'z ichiga oladi: daraxtni qurish va uni kesish; boshqalari ichki tugunlarning ko'payishini oldini olish uchun o'z ishlarida ushbu bosqichlarni almashtirib turishadi.

Daraxt qurishni to'xtatish. To'xtatish qoidasini ko'rib chiqamiz. Ko'rib chiqilayotgan tugun ichki tugun bo'ladimi-yo'qmi, aniqrog'i u yana bo'linib ketadimi yoki u yakuniy tugun ekanligini aniqlash.

To'xtatish - bu daraxtlarni qurish jarayonidagi daraxtning keying qadamlarini to'xtatish kerak bo'lgan vaqt. Qoidalarni to'xtatish variantlaridan biri bu "erta to'xtash" (oldindan kesish), bu tugunni ajratishning maqsadga muvofiqligini aniqlaydi. Ushbu parametrdan foydalanishning afzalligi modelni o'qitish vaqtini qisqartirishdir. Shu bilan birga, pastroq tasniflash aniqligi xavfi mavjud. Shuning uchun "to'xtash o'rniga qirqishdan foydalanish" tavsiya etiladi (Breiman, 1984). O'rganishni to'xtatishning ikkinchi varianti daraxt chuqurligini cheklashdir. Bunday holda, qurilish belgilangan chuqurlikka yetganda tugaydi.

To'xtatishning yana bir usuli - daraxtning barg tugunlarida mavjud bo'ladigan minimal sonli misollarni belgilash. Ushbu parametr yordamida filiallar daraxtning barcha barg tugunlari toza bo'lguncha yoki belgilangan miqdordagi moslamalarni o'z ichiga olmaguncha davom etadi.

Bir qator qoidalar mavjud, ammo shuni ta'kidlash kerakki, ularning hech biri amaliy amaliy ahamiyatga ega emas, ba'zilari esa faqat ma'lum hollarda qo'llaniladi.

Daraxtni qisqartirish yoki novdalarni kamaytirish. Haddan tashqari tarvaqaylab ketgan daraxt muammosini hal qilish uning ba'zi shoxlarini qisqartirish

orqali kamaytirish. Yechim daraxti yordamida qurilgan tasnif modelining sifati ikkita asosiy xususiyat bilan tavsiflanadi: tanib olishning aniqligi va xatosi.

Tanib olish aniqligi o'quv jarayonida to'g'ri tasniflangan ob’yektlarning mashg'ulotda qatnashgan ma'lumotlar to'plamidagi ob’yektlarning umumiy soniga nisbati sifatida hisoblanadi.

Xato o'quv jarayonida noto'g'ri tasniflangan ob’yektlarning mashg'ulotda qatnashgan ma'lumotlar to'plamidagi ob’yektlarning umumiy soniga nisbati sifatida hisoblanadi. Filiallarni qisqartirish yoki ba'zi filiallarni subtree bilan almashtirish ushbu protsedura xatoning ko'payishiga olib kelmasa amalga oshirilishi kerak. Jarayon pastdan yuqoriga qarab amalga oshiriladi, ya'ni. ko'tarilmoqda. Bu to'xtatish qoidalarini ishlatishdan ko'ra ko'proq mashhur protsedura. Ba'zi novdalarni kesgandan so'ng olingan daraxtlar kesilgan deb nomlanadi.

Agar bunday kesilgan daraxt hali ham intuitiv bo'lmasa va uni tushunish qiyin bo'lsa, sinflarni tavsiflash uchun to'plamlarga birlashtirilgan qoidalar ekstraktsiyasidan foydalaning. Daraxtning ildizidan tepasiga yoki bargigacha bo'lgan har bir yo'l bitta qoidani beradi. Qoidalarning shartlari - bu daraxtning ichki tugunlarini tekshirish.

Algoritmlar. Bugungi kunda yechim daraxtlarini amalga oshiradigan juda ko'p algoritmlar mavjud: CART, C4.5, CHAID, CN2, NewId, ITrule va boshqalar.

CART algoritmi. CART (Tasniflash va regressiya daraxti) algoritmi, nomidan ko'rinib turibdiki, tasniflash va regressiya muammolarini hal qiladi. 1974-1984 yillarda to'rtta statistika fanlari professori - Leo Breiman (Berkli), Jerri Fridman (Stenford), Charlz Stoun (Berkli) va Richard Olshen (Stenford) tomonidan ishlab chiqilgan. Ma'lumotlar to'plami atributlari diskret yoki raqamli bo'lishi mumkin.

CART algoritmi ikkilik yechimlar daraxtini yaratish uchun mo'ljallangan. Ikkilik daraxtlar ikkitomonlama daraxtlar deb ham ataladi. Bunday daraxtning namunasi ma'ruza boshida ko'rib chiqildi.

CART algoritmining boshqa xususiyatlari:

- bo'lim sifatini baholash funktsiyasi;

- daraxtlarni qisqartirish mexanizmi;
- etishmayotgan qiymatlarni qayta ishlash algoritmi;
- regressiya daraxtlarini qurish.

Ikkilik daraxtning har bir tugunida, bo'linishda, faqat ikkita qismi bor, ularni qism shoxlari deb atashadi. Filialning keyingi bo'linishi ushbu filial tomonidan dastlabki ma'lumotlarning qanchasi tasvirlanganiga bog'liq. Daraxt barpo etishning har bir bosqichida tugunda hosil bo'lgan qoida berilgan misollar to'plamini ikki qismga ajratadi. Uning o'ng qismi (o'ng filial) bu to'plamning qoida bajariladigan qismidir; chap (chap filial) - qoida bajarilmaydigan narsa.

Optimal qoidani tanlash uchun ishlatiladigan bo'limning sifatini baholash funktsiyasi - Gini indeksi yuqorida tavsiflangan edi. Ushbu funktsiya tugun noaniqligini kamaytirish g'oyasiga asoslanganligini unutmang. Aytaylik, u erda tugun bor va u ikkita sinfga bo'lingan. Tugundagi maksimal noaniqlik, uni 50 ta misoldan iborat ikkita kichik guruhga ajratishda va maksimal aniqlik - 100 va 0 misollarga bo'linishda erishiladi.

Bo'linish qoidalari. Eslatib o'tamiz, CART algoritmi raqamli va toifadagi atributlar bilan ishlaydi. Har bir tugun faqat bitta atributni ajratishi mumkin. Agar atribut raqamli bo'lsa, u holda ichki tugunda xi <= c shaklidagi qoida hosil bo'ladi.Ko'pchilik hollarda c qiymati o'zgaruvchining xi o'zgaruvchisining ikkita qo'shni tartiblangan qiymatlarining arifmetik o'rtacha qiymati sifatida tanlanadi. o'quv ma'lumotlar to'plami. Agar atribut kategorik turga tegishli bo'lsa, u holda ichki tugunda xi V (xi) qoida hosil bo'ladi, bu erda V (xi) - xi o'zgaruvchining qiymatlari to'plamining o'quv to'plamidagi bo'sh bo'lmagan kichik to'plami. .

Kesish mexanizmi. Daraxtlarni kesish uchun minimal murakkablik deb ataladigan ushbu mexanizm yordamida CART algoritmi qaror daraxtlarini qurish uchun boshqa algoritmlardan tubdan farq qiladi. Ko'rib chiqilgan algoritmda Azizillo - bu "mos o'lchamdagi" daraxtni olish va eng aniq tasnifiy bahoni olish o'rtasidagi murosaga kelishning bir turi. Usul kamayadigan daraxtlar ketma-ketligini olishdan iborat, ammo hamma daraxtlar emas, balki faqat "eng yaxshi vakillar" hisobga olinadi.

O'zaro tekshirish (V-fold Cross-validation) - bu CART algoritmining eng murakkab va shu bilan birga asl qismidir. Ma'lumotlar to'plami kichik bo'lsa yoki ma'lumotlar to'plamining yozuvlari shu qadar aniq bo'lsa, ma'lumotlar to'plamini o'quv va test to'plamlariga ajratish mumkin bo'lmagan holda, bu yakuniy daraxtni tanlash usulidir. Shunday qilib, CART algoritmining asosiy tavsiflari quyidagilardan iborat: ikkilik bo'linish, bo'linish mezonlari - Gini indeksi, minimal murakkablikdagi daraxtlarni kesish va V marta o'zaro faoliyatni tasdiqlash algoritmlari, "daraxtni o'stirib, keyin kamaytirish" printsipi, yuqori qurilish tezligi, etishmayotgan qiymatlarni qayta ishlash.

Algoritm C4.5. Algoritm C4.5 tugunda cheksiz ko'p shoxlari bo'lgan yechim daraxtini quradi. Ushbu algoritm faqat diskret qaram atribut bilan ishlashi mumkin va shuning uchun faqat tasniflash muammolarini hal qilishi mumkin. C4.5 tasnif daraxtlarini qurish uchun eng mashhur va keng qo'llaniladigan algoritmlardan biri hisoblanadi.

C4.5 algoritmining ishlashi uchun quyidagi talablar bajarilishi kerak:

- Ma'lumotlar to'plamining har bir yozuvi oldindan belgilangan sinflardan biri bilan bog'langan bo'lishi kerak, ya'ni. ma'lumotlar to'plamining atributlaridan biri sinf yorlig'i bo'lishi kerak.
- Sinflar alohida bo'lishi kerak. Har bir misol sinflarning biriga alohida murojaat qilishi kerak.
- Sinflar soni o'rganilayotgan ma'lumotlar to'plamidagi yozuvlar sonidan sezilarli darajada kam bo'lishi kerak.

Algoritmning so'nggi versiyasi C4.8 algoritmi Weka vositasida J4.8 (Java) sifatida amalga oshiriladi. Usulni tijoratda amalga oshirish: Avstraliya, RuleQuest tomonidan ishlab chiqilgan C5.0.

C4.5 algoritmi juda katta va aralash ma'lumotlar to'plamlarida sekin bajariladi.

Yechim daraxtlarini qurish uchun ikkita taniqli algoritmni ko'rib chiqdik, CART va C4.5. Ikkala algoritm ham mustahkam, ya'ni kata ma'lumotlarga nisbatan chidamli. Qaror daraxtlarini qurish algoritmlari quyidagi xususiyatlarga ko'ra farqlanadi:

- bo'linish turi - ikkilik (binary), ko'p (multi-way)

- bo'linish mezonlari - entropiya, Gini va boshqalar
- etishmayotgan qiymatlarni boshqarish qobiliyati
- daraxt shoxlarini qisqartirish yoki ularni kesishish protsedurasi
- daraxtlardan qoidalarni chiqarib olish qobiliyati.

Daraxtlarni qurish bo'yicha biron bir algoritmni eng yaxshi yoki mukammal deb hisoblash mumkin emas; ma'lum bir algoritmdan foydalanish maqsadga muvofiqligini tasdiqlash tajriba orqali tasdiqlanishi kerak.

Yangi o'lchovli algoritmlarni ishlab chiqish. Yechim daraxtlarini qurish algoritmlariga qo'yilgan eng jiddiy talab - bu miqyosi, ya'ni. algoritmda o'lchovli ma'lumotlarga kirish usuli bo'lishi kerak. Bir qator yangi o'lchovli algoritmlar ishlab chiqildi, ular orasida Sprint algoritmi Jon Shafer va uning hamkasblari tomonidan taklif qilingan. Ma'ruzada muhokama qilingan CART algoritmining kengaytiriladigan versiyasi bo'lgan Sprint operativ xotira hajmiga minimal talablarni qo'yadi.

XULOSA. Mashg`ulotda biz yechimlar daraxtini ko'rib chiqdik, bu qisqacha ob’yektlarning ma'lum bir sinfga mansubligini taxmin qilish yoki raqamli o'zgaruvchilar qiymatlarini bashorat qilishning ierarxik, moslashuvchan vositasi sifatida ta'riflanishi mumkin. Ko'rib chiqilayotgan qarorlar daraxti usulining ishlash sifati ham algoritm tanlashga, ham o'rganilgan ma'lumotlar to'plamiga bog'liq. Ushbu uslubning barcha afzalliklariga qaramay, sifatli modelni yaratish uchun qaram va mustaqil o'zgaruvchilar o'rtasidagi munosabatlarning mohiyatini tushunish va yetarli ma'lumot to'plamini tayyorlash kerakligini esdan chiqarmaslik kerak.

Nazorat savollari:

1. Algoritm nima?
2. Yechim daraxti deganda nimani tushunasiz?
3. Yechim daraxtlari orqali har doim yechimga erishish mumkinmi?
4. Yechimlar daraxtining qanday afzalliklari mavjud?

10-MAVZU

OPOR VEKTORLAR USULI. BASESLI KLASSIFIKATSIYA

Reja:

1. **Vektorli metodlarni qo'llab-quvvatlash**
2. **Basesli klassifikatsiya**

Mashg`ulot maqsadi*: Qo'llab-quvvatlash vektori mashinasining asosiy g'oyalari, "eng yaqin qo'shni" va Bases tasnifi. Ushbu usullarning afzalliklari va kamchiliklari ko'rib chiqiladi.*

Tayanch iboralar: *qo'llab-quvvatlash vektori, presedent, tayanch, vektorli mashina, ikkilik tasnif, tekislik, tekislik, ob'yekt, dasturiy ta'minot, to'g'ri chiziq, to'plamlar, sinflar soni, qidirish, giperplan, funktsiya, qiymat, kutilayotgan xavf, empirik tavakkal, minimallashtirish, makon, operator yadrolari, o'lchov, komponent, birgalikda ma'lumotlar, CASE, mulohaza yuritish, CBR, chiqish, mavzu maydoni, o'quv namunasi, mahalla, algoritm, grafik, chiqish, chiqishni o'rtacha hisoblash, ehtimollik, parametr, murosaga kelish, o'zaro tekshirish, segmentlar, aniqlik, foiz , namuna hajmi, dasturiy ta'minot, xulosa chiqarish, intranet, ma'lumotlar qazib olish, nuqta, javob, vositalar, namunalarni tanib olish, dastgoh, alternativa, modellashtirish, statistika, bilimlarni rasmiylashtirish, tushuncha, yondashuv, o'zaro mustaqil, qayta tayyorlash, oqim, yozish.*

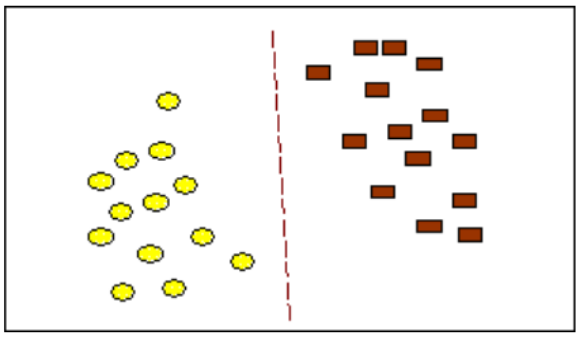
Oldingi mavzularda biz chiziqli regressiya va qaror daraxtlari kabi tasniflash va bashorat qilish usullarini ko'rib chiqdik; ushbu ma'ruzada biz ushbu guruhning usullari bilan tanishishni davom ettiramiz va quyidagilarni ko'rib chiqamiz: qo'llab-quvvatlash vektor mashinasi, eng yaqin qo'shni usuli (foydalanish holatlarida fikrlash usuli) va Bayes tasnifi.

1. Vektorli metodlarni qo'llab-quvvatlash

Vektorli kompyuterni qo'llab-quvvatlash (SVM) chegara usullari guruhiga kiradi. U ko'lam chegaralaridan foydalangan holda sinflarni belgilaydi.

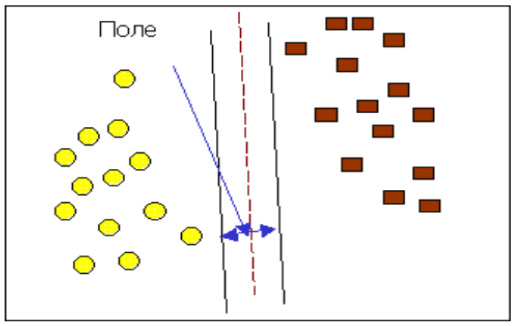
Ushbu usul ikkilik tasniflash muammolarini hal qilish uchun ishlatiladi. Usul echimlar tekisliklari tushunchasiga asoslangan. Yotiq tekisligi turli sinflarga ega bo'lgan ob'yektlarni ajratib turadi. 10.1.1-shaklda ikki turdagi ob'yektlarni o'z ichiga

olgan misol keltirilgan. Ajratuvchi chiziq chegarani belgilaydi, uning o'ng tomonida - jigarrang (jigarrang) turdagi barcha narsalar, chap tomonida esa sariq (sariq) turdagi narsalar. O'ngga tushgan yangi ob’yekt, jigarrang sinf ob’yekti sifatida, yoki ajratish chizig'ining chap tomonida joylashgan bo'lsa, sariq sinf ob’yekti sifatida tasniflanadi. Bunday holda, har bir ob’yekt ikki o'lchov bilan tavsiflanadi.



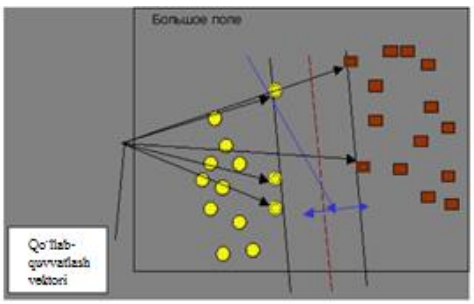
Shakl: 10.1.1.Sinflarni to'g'ri chiziq bilan ajratish

Qo'llab-quvvatlovchi vektorli mashinalarning maqsadi - ob’yektlarning ikkita to'plamini ajratib turadigan tekislikni topish; bunday tekislik shakl. 10.1.2. Ushbu rasmda namunalar to'plami ikki sinfga bo'lingan: sariq rangli narsalar A sinfiga, jigarrang narsalar B sinfiga tegishli.



Shakl: 10.1.2. Qo'llab-quvvatlash vektorlarining ta'rifiga asosan.

Usul ikki sinf o'rtasidagi chegaralarni topadi, ya'ni. qo'llab-quvvatlash vektorlari; ular shakl. 10.1.3.

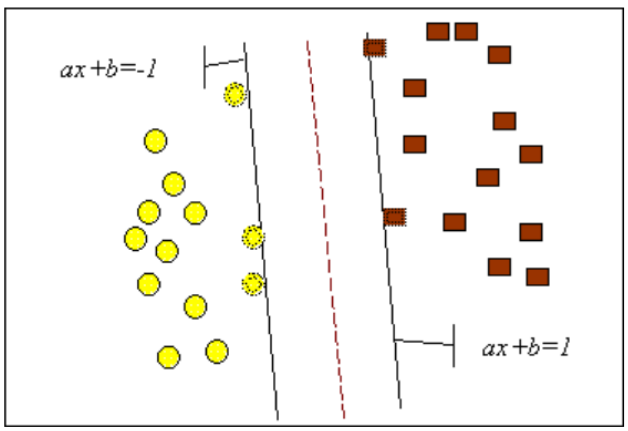


Shakl: 10.1.3. Qo'llab-quvvatlash vektorlari

Qo'llab-quvvatlash vektorlari - bu mintaqalar chegaralarida joylashgan to'plam ob’yekti. Agar chegaralar orasidagi maydon bo'sh bo'lsa, tasnif yaxshi deb hisoblanadi. Shakl. 10.1.3 Ushbu to'plamni qo'llab-quvvatlaydigan beshta vektor ko'rsatilgan.

Chiziqli SVM. Qo'llab-quvvatlash vektori mashinasi yordamida ikkilik tasniflash muammosini hal qilish ma'lumotlar bazasini to'g'ri ravishda ikki sinfga ajratadigan ba'zi bir chiziqli funktsiyalarni topishdir. Sinflar soni ikkita bo'lgan tasniflash muammosini ko'rib chiqing. Muammoni bir sinf vektorlari uchun noldan kichik, boshqa sinf vektorlari uchun noldan katta qiymatlarni qabul qiladigan f (x) funktsiyani qidirish sifatida shakllantirish mumkin. Muammoni hal qilish uchun dastlabki ma'lumotlar sifatida, ya'ni. f (x) tasniflash funktsiyasini izlash, ularning sinflardan biriga mansubligi ma'lum bo'lgan kosmik vektorlarning o'quv to'plami berilgan. Funktsiyalarni tasniflash oilasini f (x) funktsiya nuqtai nazaridan tavsiflash mumkin. Giperplane a vektori va b qiymati bilan aniqlanadi, ya'ni. f (x) = ax + b. Ushbu muammoning echimi shakl. 10.1.4.

Muammoni hal qilish natijasida, ya'ni. SVM modelini qurish uchun bir sinf vektorlari uchun noldan kichik, boshqa sinf vektorlari uchun noldan katta qiymatlarni oladigan funktsiya topiladi. Har bir yangi ob’yekt uchun salbiy yoki ijobiy qiymat ob’yekt sinflardan biriga tegishli ekanligini aniqlaydi.



Shakl: 10.1.4. Chiziqli SVM

Eng yaxshi tasniflash funktsiyasi - bu kutilgan xavf minimaldir. Bu holda kutilayotgan xavf tushunchasi tasnif xatolarining kutilayotgan darajasini anglatadi.

Qurilgan modelning kutilayotgan xatolik darajasini to'g'ridan-to'g'ri baholash mumkin emas; bu empirik tavakkal tushunchasi yordamida amalga oshirilishi mumkin. Ammo shuni yodda tutish kerakki, ikkinchisini minimallashtirish har doim ham kutilgan xavfni minimallashtirishga olib kelmaydi. Nisbatan kichik o'quv ma'lumotlari to'plamlari bilan ishlashda buni yodda tuting.

Empirik tavakkal - bu mashg'ulotlar to'plamidagi tasnif xatolarining darajasi.

Shunday qilib, chiziqli ajratilgan ma'lumotlar uchun qo'llab-quvvatlovchi vektorli mashina tomonidan muammoni hal qilish natijasida biz kutilgan xavfning yuqori bahosini minimallashtiradigan tasniflash funktsiyasini olamiz.

Ko'rib chiqilayotgan usul bo'yicha tasniflash muammolarini hal qilish bilan bog'liq muammolardan biri bu ikki sinf o'rtasida chiziqli chegarani topish har doim ham oson bo'lmasligi.

Bunday holatlarda variantlardan biri bu o'lchamni oshirishdir, ya'ni. ma'lumotlarni tekislikdan uch o'lchovli kosmosga uzatish, bu erda ko'plab namunalarni ikkita sinfga ideal tarzda ajratadigan tekislik qurish mumkin. Bunday holda, qo'llab-quvvatlash vektorlari har ikkala sinfning haddan tashqari ob’yektlari bo'ladi. Shunday qilib, yadro operatori deb nomlangan va qo'shimcha o'lchamlarni qo'shib, sinflar orasidagi chegaralar giperplanalar ko'rinishida topiladi. Shunga qaramay, SVM modelini qurishning murakkabligi shundan iboratki, fazoning kattaligi qanchalik baland bo'lsa, u bilan ishlash shunchalik qiyin bo'ladi. Yuqori o'lchovli ma'lumotlar bilan ishlashning variantlaridan biri bu eng muhim tarkibiy qismlarni aniqlash uchun avval ma'lumotlarning o'lchamlarini kamaytirishning ba'zi usullaridan foydalanish va keyin qo'llab-quvvatlash vektorli mashinadan foydalanishdir.

Boshqa usullar singari, SVM usuli ham ushbu usulni tanlashda e'tiborga olinishi kerak bo'lgan kuchli va zaif tomonlarga ega. Ushbu usulning nochorligi shundaki, tasniflash uchun barcha namunalar to'plamidan emas, balki ularning chegaralarida joylashgan kichik qismidan foydalaniladi. Usulning afzalligi shundaki, ko'plab boshqa usullardan farqli o'laroq, qo'llab-quvvatlash vektorlari tasnifi uchun kichik ma'lumotlar to'plami etarli. Sinov to'plamida qurilgan

modelning to'g'ri ishlashi bilan ushbu usuldan real ma'lumotlarda foydalanish mumkin.

Qo'llab-quvvatlash vektori mashinasiga imkon beradi:

- kutilayotgan tavakkalchilikning minimal yuqori darajasi bilan tasniflash funktsiyasini oling (tasniflash xatosi darajasi);
- soddaligini samaradorlik bilan birlashtirib, chiziqli bo'lmagan ajratilgan ma'lumotlar bilan ishlash uchun chiziqli tasniflagichdan foydalaning.

2. Basesli klassifikatsiyasi

Muqobil nomlar: Bases modellashtirish, Bases statistikasi, Bases tarmog'i usuli.

Dastlab ekspert tizimlarida ekspert bilimlarini rasmiylashtirish uchun Bases tasnifi ishlatilgan, endi Bases tasnifi ham Data Mining usullaridan biri sifatida foydalanilmoqda. Sodda tasniflash yoki 122oda Bases tarmoqlari yordamida usulning eng oddiy versiyasidir. Ushbu yondashuv tasniflash muammolarini hal qiladi, usulning natijasi “shaffof” modellardir.

“Sodda” tasniflash – bu tasnifning etarlicha shaffof va tushunarli usuli. U “122oda” deb nomlanadi, chunki u xususiyatlarning o’zaro mustaqilligini taxmin qilishga asoslanadi.

Sodda tasniflash xususiyatlari:

1. Barcha o’zgaruvchilardan foydalanish va ular orasidagi barcha bog’liqliklarni aniqlash.
2. O’zgaruvchilar haqida ikkita taxminga ega bo’lish:

- barcha o’zgaruvchilar bir xil darajada muhimdir;
- barcha o’zgaruvchilar 122oda122vall jihatdan mustaqil, ya’ni. Bitta o’zgaruvchining qiymati boshqasining qiymati haqida hech narsa demaydi.

Ko’pgina boshqa tasniflash usullari ob’yektning u yoki bu sinfga tegishli bo’lish ehtimoli tasniflash boshlanishidan oldin bir xil deb taxmin qiladi; ammo bu har doim ham to’g’ri emas.

Aytaylik, ma’lumotlarning ma’lum bir qismi ma’lum bir sinfga tegishli ekanligini bilasiz. Savol tug’iladi, tasniflash modelini tuzishda ushbu ma’lumotdan foydalanishimiz mumkinmi? Ob’yektlarni tasniflashda yordam berish uchun ushbu

oldingi bilimlardan foydalanishning ko’plab haqiqiy misollari mavjud. Tibbiy amaliyotdan odatiy misol. Agar shifokor bemorning test natijalarini qo’shimcha tadqiqotlar uchun yuboradigan bo’lsa, u bemorni ma’lum bir sinfga ajratadi. Ushbu ma’lumotni qanday qo’llash mumkin? Tasniflash modelini tuzishda uni qo’shimcha ma’lumotlar sifatida ishlatishimiz mumkin.

Data Mining usuli sifatida Bases tarmoqlarining afzalliklari qayd etilgan:

- model barcha o’zgaruvchilar o’rtasidagi bog’liqlikni belgilaydi, bu ba’zi bir o’zgaruvchilarning qiymatlari noma’lum bo’lgan vaziyatlarni boshqarishni osonlashtiradi;
- Bases tarmoqlarini talqin qilish juda oson va bashoratli modellashtirish bosqichida oson tahlil qilishga imkon beradi;
- Bases usuli sizga ma’lumotlardan kelib chiqadigan naqshlarni va, masalan, aniq shaklda olingan ekspert bilimlarini tabiiy ravishda birlashtirishga imkon beradi;
- Bases tarmog’idan foydalanish ortiqcha fitnadan, ya’ni modelning haddan tashqari murakkablashuvidan qochadi, bu ko’plab usullarning zaif tomoni (masalan, qaror daraxtlari va neyron tarmoqlari).

Baseslarning 123oda yondashuvi quyidagi kamchiliklarga ega:

- ko’paytirish shartli ehtimolliklar faqat barcha kiritilgan o’zgaruvchilar haqiqatan ham 123oda123vall jihatdan mustaqil bo’lganda to’g’ri bo’ladi; 123oda123vall mustaqillik sharti bajarilmasa, ushbu usul ko’pincha juda yaxshi natijalarni ko’rsatsa-da, nazariy jihatdan ushbu vaziyatni Bayes tarmoqlarini tayyorlashga asoslangan ancha murakkab usullar bilan hal qilish kerak;
- uzluksiz o’zgaruvchilarni to’g’ridan-to’g’ri qayta ishlashning iloji yo’q – atributlar diskret bo’lishi uchun ularni 123oda123vallic shkalaga aylantirish kerak; ammo, bunday transformatsiyalar ba’zan muhim naqshlarning yo’qolishiga olib kelishi mumkin;
- 123oda Bayes yondashuvida tasniflash natijasiga faqat kirish o’zgaruvchilarining individual qiymatlari ta’sir qiladi; bu erda turli xil

atributlar qiymatlari juftliklari yoki uchliklarining qo’shma ta’siri hisobga olinmaydi. Bu tasniflash modelining sifatini taxminiy aniqligi jihatidan yaxshilashi mumkin, ammo sinovdan o’tgan variantlar sonini ko’paytiradi.

Bases tasnifi amalda keng qo’llanilishini topdi.

Bases usuli orqali so’zlarni filtrlash. Yaqinda spamni shaxsiy filtrlash uchun Bases tasnifi taklif qilindi. Birinchi filtr Pol Grem tomonidan ishlab chiqilgan. Algoritm ikkita talabni bajarishni talab qiladi. Birinchi talab, tasniflangan ob'yekt yetarli miqdordagi xususiyatlarga ega bo'lishi kerak. Foydalanuvchi xatlaridagi barcha so'zlar buni juda yaxshi qondiradi, juda qisqa va juda kamdan-kam hollarda. Ikkinchi talab - "spam - spam emas" to'plamini doimiy ravishda qayta tayyorlash va to'ldirish. Bunday shartlar mahalliy pochta mijozlarida juda yaxshi ishlaydi, chunki oxirgi mijozdan "spam bo'lmagan" oqim doimiy ravishda o'zgarib turadi va agar u o'zgarsa, tezda bo'lmaydi. Shu bilan birga, serverning barcha mijozlari uchun "spam bo'lmagan" oqimini aniq aniqlash juda qiyin, chunki bitta mijoz uchun spam bo'lgan xabar boshqasiga spam emas. Lug'at juda katta bo'lib chiqadi, spam va "spam emas" deb aniq ajratish mumkin emas, natijada tasniflash sifati, bu holda harflarni filtrlash muammosining echimi sezilarli darajada kamayadi.

Nazorat savollari:

1. Bases usulini tushuntirib bering.
2. Sodda tasniflash usullariga misol keltiring.
3. Qanday tasniflash usullari mavjud?
4. Vektor deganda nimani tushunasiz?
5. Klassifikatsiya nima?

11-MAVZU

NEYRON TO`RLARI. KLASTERLI TAHLIL

Reja:

1. **Neyron to`rlari**
2. **Neyron tarmoq arxitekturasi**
3. **Neyron tarmoq modellari**

Mashg`ulot maqsadi: *Ma'ruzada neyron tarmoqlari usuli tasvirlangan. Elementlar va arxitektura, o'quv jarayoni va asab tarmog'ini qayta tayyorlash fenomeni ko'rib chiqiladi. Perseptron kabi bunday neyron tarmoq modeli tasvirlangan. Nerv tarmoqlari apparati yordamida muammoni echishga misol keltirilgan.*

Tayanch iboralar: *neyron tarmoq, yo'naltirilgan grafik, avtomatlashtirish, identifikatsiya qilish, tanib olish, o'z-o'zini o'rganish tizimi, qatlam, ma'lumot, ma'lumotlar, nazoratsiz o'rganish, o'zini o'zi tashkil etuvchi xarita, ma'lumotlar bazasi, ta'rif, sun'iy neyron, neyron, funktsiya, chiziqli bo'lmagan transformator, funktsiyani faollashtirish, chiqish, filial nuqtasi, sinaps, akson, aloqa, faollashtirish funktsiyasi, dasturiy ta'minot, qatlamli asab tarmog'i, pertseptron, asosiy funktsiyalar, kognitron, neokognitron, navbat, tarmoq, kirish, chiqish, yashirin, kirish neyroni, kirish, neyron, yashirin neyron, yashirin, chiqish neyron, chiqish, ichki parametrlar, algoritm, davr, iteratsiya, to'plamlar, o'quv namunasi, chiziqli model, tahlilchi, farq, o'rganish xatosi, xato funktsiyasi, maqsad funktsiyasi, qayta tayyorlash, prognoz aniqligi, bo'linish, MLP, orqaga tarjima usuli xatolar, tarqatish, qoldiq, dasturiy ta'minot, kredit, maydon, foydalanuvchi, ko'rsatuvchi, jadval, grafik, tahlil, qiymat, asboblar qutisi, o'zgaruvchi, tarmoq o qardoshlik bilan tarqalish, sintaksis, massiv, SI, eng kichik kvadratlar, to'r, SSE, MAT, qurilish.*

1. **Neyron to`rlari.**

2. Neyron tarmoqlari g'oyasi sun'iy intellekt nazariyasi doirasida biologik asab tizimining o'rganish va xatolarni tuzatish qobiliyatini taqlid qilishga urinishlar natijasida paydo bo'ldi.

Neyron tarmoqlari - bu miyadagi biologik neyron tarmoqlarining modellari bo'lib, unda neyronlar nisbatan sodda, ko'pincha bir xil turdagi elementlar (sun'iy neyronlar) tomonidan taqlid qilinadi. Neyron tarmog'ini sun'iy neyronlar tepaliklar, sinaptik bog'lanishlar esa yoy bo'lgan og'irlikdagi ulanishlar bilan yo'naltirilgan grafik bilan ifodalash mumkin. Neyron tarmoqlari turli xil muammolarni hal qilishda keng qo'llaniladi. Neyron tarmoqlarini qo'llash sohalari orasida naqshlarni tanib olish jarayonlarini avtomatlashtirish, prognozlash, moslashuvchan boshqarish, ekspert tizimlarini yaratish, assotsiativ xotirani tashkil etish, analog va raqamli signallarni qayta ishlash, elektron sxemalar va tizimlarni sintez qilish va identifikatsiyalash kiradi.

Neyron tarmoqlaridan foydalangan holda siz, masalan, mahsulot sotish hajmini, fond bozori ko'rsatkichlarini bashorat qilishingiz, signallarni aniqlashni amalga oshirishingiz va o'z-o'zini o'rganish tizimlarini loyihalashingiz mumkin.
Neyron tarmoq modellari dasturiy ta'minot va apparatning bajarilishi bo'lishi mumkin. Biz birinchi turdagi tarmoqlarni ko'rib chiqamiz.

Oddiy qilib aytganda, qatlamli asab tarmog'i bu qatlamlarni tashkil etuvchi neyronlarning to'plamidir. Har bir qatlamda neyronlar bir-birlari bilan hech qanday bog'liq emas, lekin ular oldingi va keyingi qatlamlarning neyronlari bilan bog'langan. Axborot birinchi qavatdan ikkinchi qavatgacha, ikkinchidan uchinchi qavatgacha va hokazo.

Neyron tarmoqlari yordamida hal qilingan Data Mining vazifalari orasida biz quyidagilarni ko'rib chiqamiz:

- Klassifikatsiya (nazorat ostida o'rganish). Klassifikatsiya vazifalarining namunalari: matnni aniqlash, nutqni aniqlash, shaxsni aniqlash.
- Bashorat qilish. Neyron tarmog'i uchun bashorat qilish muammosi quyidagicha shakllantirilishi mumkin: kirish qiymatlarining cheklangan

to'plami tomonidan berilgan funktsiyaning eng yaxshi yaqinlashishini toping (o'quv misollari). Masalan, neyron tarmoqlar etishmayotgan qiymatlarni tiklash muammosini hal qilishga imkon beradi.

- Klasterlash (nazoratsiz o'rganish). Ma'lumotlarning hajmini kamaytirish orqali ma'lumotni siqish muammosi klasterlash muammosiga misol bo'lishi mumkin. Klasterlash muammolari, masalan, Kohonen xaritalarini o'z-o'zini tartibga solish orqali hal qilinadi. Ushbu tarmoqlarga alohida ma'ruza bag'ishlanadi.

Keling, neyron tarmoqlardan foydalanish mumkin bo'lgan uchta vazifa misolini ko'rib chiqamiz.

Tibbiy diagnostika. Bemorning ahvolining turli ko'rsatkichlarini kuzatish jarayonida ma'lumotlar bazasi to'plandi. Asoratlar xavfi kuzatilgan o'zgaruvchilarning murakkab chiziqli bo'lmagan birikmasiga mos kelishi mumkin, bu neyron tarmoqlarni modellashtirish yordamida aniqlanadi.

Kompaniya ish faoliyatini prognoz qilish (sotish). Tashkilot faoliyati to'g'risidagi retrospektiv ma'lumotlarga asoslanib, kelgusi davrlar uchun savdo hajmini aniqlash mumkin.

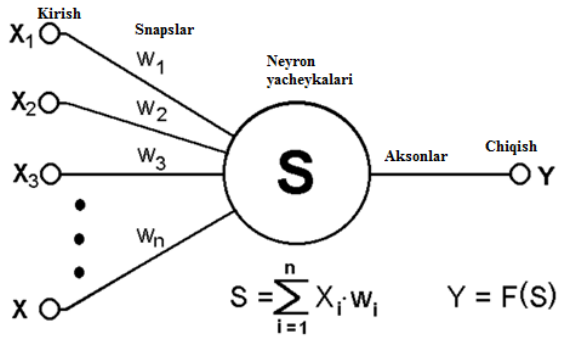
Kredit berish. Bankning mijozlar ma'lumotlar bazasidan foydalangan holda, neyron tarmoqlaridan foydalanib, potentsial "defoltlar" guruhiga kiruvchi mijozlar guruhini tashkil qilish mumkin.

Neyron tarmoqlarining elementlari

Sun'iy neyron (rasmiy neyron) - bu biologik neyronning ba'zi funktsiyalarini simulyatsiya qiladigan sun'iy neyron tarmoqlarining elementi.

Sun'iy neyronning asosiy vazifasi uning kirishiga kelgan signallarga qarab chiqish signalini hosil qilishdir. Eng keng tarqalgan konfiguratsiyada kirish signallari moslashuvchan biriktiruvchi tomonidan qayta ishlanadi, so'ngra qo'shimchaning chiqishi chiziqli bo'lmagan konvertorga o'tadi, u erda u faollashtirish funktsiyasi bilan aylantiriladi va natija chiqishga (tarmoq nuqtasiga) beriladi.

Sun'iy neyronning umumiy ko'rinishi shakl. 11.1.1.



Shakl: 11.1.1. Sun'iy neyron

Neyron hozirgi holati bilan ajralib turadi va boshqa neyronlarning chiqishi bilan bog'langan bir yo'nalishli kirish aloqalari - sinapslar guruhiga ega.

Neyronda akson mavjud - ma'lum bir neyronning chiqish aloqasi, undan quyidagi neyronlarning sinapslariga signal (qo'zg'alish yoki inhibisyon) yuboriladi.

Har bir sinaps sinaptik ulanish kattaligi (uning og'irligi wi) bilan tavsiflanadi.

Neyronning hozirgi holati uning kirishlarining tortilgan yig'indisi sifatida aniqlanadi:

$$s = \sum_{i=1}^{n} x_i \cdot w_i$$

Neyronning chiqishi uning holatiga bog'liq:

$$y = f(s).$$

Faollashtirish funktsiyasi, shuningdek xarakterli funktsiya deb ataladi, bu rasmiy neyronning chiqishini hisoblaydigan chiziqli bo'lmagan funktsiya.

Tez-tez ishlatiladigan faollashtirish funktsiyalari:

- Qattiq jegaraviy funktsiyasi.
- Chiziqli chegara.
- Sigmoidal funktsiya.

Faollashtirish funktsiyasini tanlash qo'yilayotgan vazifaning o'ziga xos xususiyatlari yoki ba'zi bir o'rganish algoritmlari tomonidan qo'yilgan cheklovlar bilan belgilanadi.

Chiziqli bo'lmagan konvertor - bu neyronning hozirgi holatini (moslashuvchan qo'shimchaning chiqish signali) ba'zi bir chiziqli bo'lmagan qonunga ko'ra (faollashtirish funktsiyasi) neyronning chiqish signaliga aylantiradigan sun'iy neyronning elementidir.

Tugun nuqtasi (chiqish) - bu chiqish signalini bir nechta manzillarga yuboradigan va bitta kirish va bir nechta chiqishga ega bo'lgan rasmiy neyronning elementi.

Tugun nuqtasining kiritilishi odatda chiziqli bo'lmagan transduserning chiqishi bo'lib, u keyinchalik boshqa neyronlarning kirishiga yuboriladi.

3. Neyron tarmoq arxitekturasi

Neyron tarmoqlari sinxron va asinxron bo'lishi mumkin.

Sinxron asab tizimlarida bir vaqtning o'zida faqat bitta neyron o'z holatini o'zgartiradi. Asenkronlarda holat bir vaqtning o'zida neyronlarning butun bir guruhida, qoida tariqasida, butun qatlamda o'zgaradi. Ikkita asosiy me'morchilikni ajratish mumkin - qatlamli va to'liq bog'langan tarmoqlar. Qatlamli tarmoqlarning kaliti - bu qatlam tushunchasi.

Qatlam - kirishlari bir xil umumiy signal bilan bog`liq bir yoki bir nechta neyron.

Qatlamli neyron tarmoqlari - bu neyronlar alohida guruhlarga (qatlamlarga) bo'linadigan, shuning uchun axborotni qayta ishlash qatlamlarda amalga oshiriladigan nerv tarmoqlari.

Qatlamli tarmoqlarda i-qavatning neyronlari kirish signallarini qabul qiladi, ularni o'zgartiradi va (i + 1) qavat neyronlariga tarmoq uchlari orqali uzatadi. Va tarjimon va foydalanuvchi uchun chiqish signallarini ta'minlovchi k-chi qatlamgacha. Har bir qatlamdagi neyronlarning soni boshqa qatlamlardagi neyronlarning soni bilan bog'liq emas, u o'zboshimchalik bilan bo'lishi mumkin. Bitta qatlam ichida ma'lumotlar parallel ravishda qayta ishlanadi va butun tarmoq bo'ylab qayta ishlash ketma-ket - qatlamdan qatlamgacha amalga oshiriladi. Qatlamli neyron tarmoqlariga, masalan, ko'p qavatli pertseptronlar, radial asosli funktsiyalar tarmoqlari, kognitron, neokognitron va assotsiativ xotira tarmoqlari

kiradi. Biroq, signal har doim ham qatlamdagi barcha neyronlarga qo'llanilmaydi. Masalan, kognitronda hozirgi qatlamning har bir neyroni signallarni faqat oldingi qatlamda unga yaqin bo'lgan neyronlardan oladi.

Qatlamli tarmoqlar, o'z navbatida, bir qavatli va ko'p qavatli bo'lishi mumkin [46].

Bir qavatli tarmoq - bitta qatlamdan iborat bo'lgan tarmoq.

Ko'p qatlamli tarmoq - bir necha qatlamlarga ega bo'lgan tarmoq.

Ko'p qavatli tarmoqda birinchi qavat kirish qavat, keyingilari ichki yoki yashirin, oxirgi qavat esa chiqish qavat deyiladi. Shunday qilib, oraliq qatlamlar ko'p qavatli asab tarmog'idagi barcha qatlamlardir, faqat kirish va chiqishdan tashqari.

Tarmoqning kirish qatlami kirish ma'lumotlari bilan chiqishni, chiqish - chiqish bilan aloqani amalga oshiradi. Shunday qilib, neyronlar kirish, chiqish va yashirin bo'lishi mumkin. Kirish qatlami ma'lumotlarni qabul qiladigan va uni tarmoqning yashirin qatlami neyronlarining kirishiga tarqatadigan kirish neyronlaridan tashkil etilgan.

Yashirin neyron - bu asab tarmog'ining yashirin qatlamida joylashgan neyron.

Tarmoqning chiqish qatlami tashkil qilingan **chiqish neyronlari** neyron tarmoq faoliyati natijalarini keltirib chiqaradi.

To'liq ulangan tarmoqlarda har bir neyron o'zining signal signalini o'zi bilan birga qolgan neyronlarga uzatadi. Tarmoqning chiqish signallari tarmoqning bir necha soatlik tsikllaridan keyin neyronlarning chiqish signallarining barchasi yoki bir nechtasi bo'lishi mumkin. Barcha kirish signallari barcha neyronlarga o'tadi.

Neyron tarmoqlari bo'yicha trening

Neyron tarmog'ini ishlatishdan oldin uni o'rgatish kerak. Neyron tarmog'ini o'qitish jarayoni ma'lum bir vazifa uchun uning ichki parametrlarini sozlashdan iborat. Neyron tarmog'ining algoritmi takrorlanuvchi bo'lib, uning bosqichlari davrlar yoki tsikllar deb ataladi.

Davr - bu o'quv jarayonidagi bitta takrorlanish bo'lib, u o'quv majmuasidagi barcha misollarni taqdim etishni va, ehtimol, nazorat to'plamida o'qitish sifatini tekshirishni o'z ichiga oladi.

O'quv jarayoni o'quv namunasi bo'yicha amalga oshiriladi. O'quv namunasi ma'lumotlar bazasidan kirish qiymatlarini va ularga mos keladigan chiqish qiymatlarini o'z ichiga oladi. Trening davomida asab tarmog'i kirish maydonlarining ba'zi bog'liqligini topadi. Shunday qilib, biz savolga duch kelmoqdamiz - biz qanday kirish maydonlarini (xususiyatlarini) ishlatishimiz kerak. Dastlab, tanlov evristik usulda amalga oshiriladi, so'ngra kirishlar sonini o'zgartirish mumkin. Ma'lumotlar to'plamidagi kuzatuvlar soni haqidagi savol biroz qiyin bo'lishi mumkin. Kerakli kuzatuvlar soni va tarmoq hajmi o'rtasidagi bog'liqlikni tavsiflovchi ba'zi qoidalar mavjud bo'lsa ham, ularning to'g'riligi isbotlanmagan. Kerakli kuzatuvlar soni hal qilinayotgan muammoning murakkabligiga bog'liq. Xususiyatlar sonining ko'payishi bilan kuzatuvlar soni chiziqsiz ravishda ko'payadi, bu muammo "o'lchov la'nati" deb nomlanadi. Agar ma'lumot etarli bo'lmasa, chiziqli modeldan foydalanish tavsiya etiladi. Analitik tarmoqdagi qatlamlar sonini va har bir qatlamdagi neyronlarning sonini aniqlashi kerak. Keyinchalik, siz qaror xatoligini minimallashtirishi mumkin bo'lgan og'irlik va noto'g'ri tomonlarning bunday qiymatlarini belgilashingiz kerak. Og'irliklar va xatoliklar avtomatik ravishda kerakli xato va o'qish xatosi deb nomlangan chiqish signali orasidagi farqni minimallashtirish uchun o'rnatiladi. Tuzilgan neyron tarmoq uchun o'rganish xatosi chiqish va maqsad (kerakli) qiymatlarini taqqoslash yo'li bilan hisoblanadi. Xato funktsiyasi olingan farqlardan hosil bo'ladi.

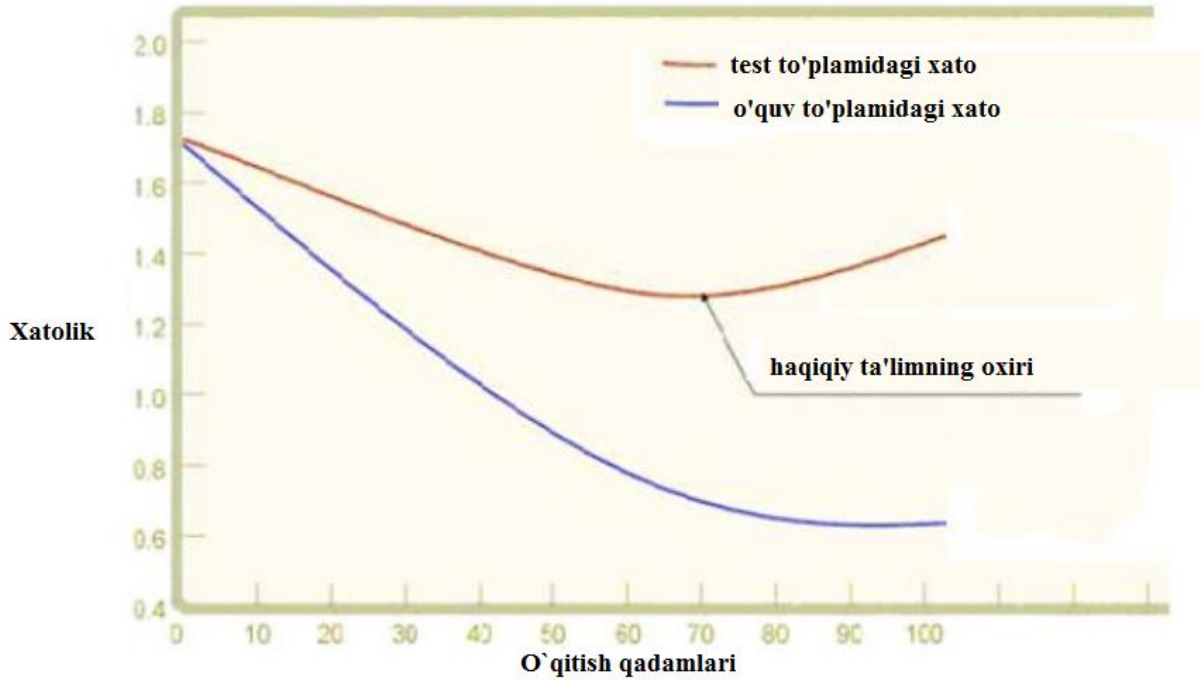
Xato funktsiyasi ob’yektiv funktsiya bo'lib, uni neyron tarmoqni boshqariladigan o'rganish jarayonida minimallashtirish kerak.

Xato funktsiyasidan foydalangan holda siz mashg'ulotlar davomida neyron tarmoq sifatini baholashingiz mumkin. Masalan, kvadratik xatolar yig'indisi ko'pincha ishlatiladi. Neyron tarmog'ini tayyorlash sifati uning oldiga qo'yilgan vazifalarni hal qilish qobiliyatini belgilaydi.

Neyron tarmog'ini qayta tayyorlash. Neyron tarmoqlarini o'rgatish paytida ko'pincha ortiqcha fitnalar muammosi deb ataladigan jiddiy qiyinchiliklar mavjud.

Haddan tashqari moslashtirish yoki o'ta yaqin moslashtirish - bu tarmoqni umumlashtirish qobiliyatini yo'qotadigan ma'lum bir o'quv misollari to'plamiga juda aniq mos keladigan neyron tarmog'i.

Haddan tashqari moslashtirish juda uzoq muddatli mashg'ulotlar, etarli miqdordagi o'quv misollari yoki asab tarmog'ining haddan tashqari murakkab tuzilishida yuz beradi. Haddan tashqari moslashish mashg'ulot (mashg'ulot) to'plamini tanlash tasodifiy ekanligi bilan bog'liq. O'rganishning dastlabki bosqichlaridan boshlab xatolar kamayadi. Keyingi bosqichlarda xatoni kamaytirish uchun (ob’yektiv funktsiya) parametrlar o'quv majmuasi xususiyatlariga moslashtiriladi. Ammo, bu holda, "sozlash" ketma-ketlikning umumiy qonunlari asosida emas, balki uning qismi - mashg'ulotning pastki qismining xususiyatlari ostida amalga oshiriladi. Bunday holda, prognozning aniqligi pasayadi. Tarmoqni ortiqcha moslashtirish bilan kurashish variantlaridan biri bu o'quv namunasini ikki to'plamga bo'lish (trening va test). Neyron tarmog'i o'quv majmuasida o'qitiladi. Qurilgan model sinov to'plamida tekshiriladi. Ushbu to'plamlar bir-birining ustiga chiqmasligi kerak. Har bir qadamda modelning parametrlari o'zgaradi, ammo maqsad funktsiyasi qiymatining doimiy pasayishi aniq o'quv majmuasida sodir bo'ladi. To'plamni ikkiga ajratganda, biz test to'plamidagi prognoz xatolarining o'zgarishini mashg'ulotlar to'plami bo'yicha kuzatuvlar bilan parallel ravishda kuzatishimiz mumkin. Ikkala to'plamda ham prognoz qilingan xatolikning ma'lum bir bosqichi kamayadi. Biroq, ma'lum bir qadamda, sinovlar to'plamidagi xatolik ko'paya boshlaydi, mashg'ulotlardagi xatolar esa kamayib boraveradi. Ushbu lahza haqiqiy yoki haqiqiy ta'limning oxiri deb hisoblanadi va qayta tayyorlash shu bilan boshlanadi. Ta'riflangan jarayon shakl. 11.2.1.



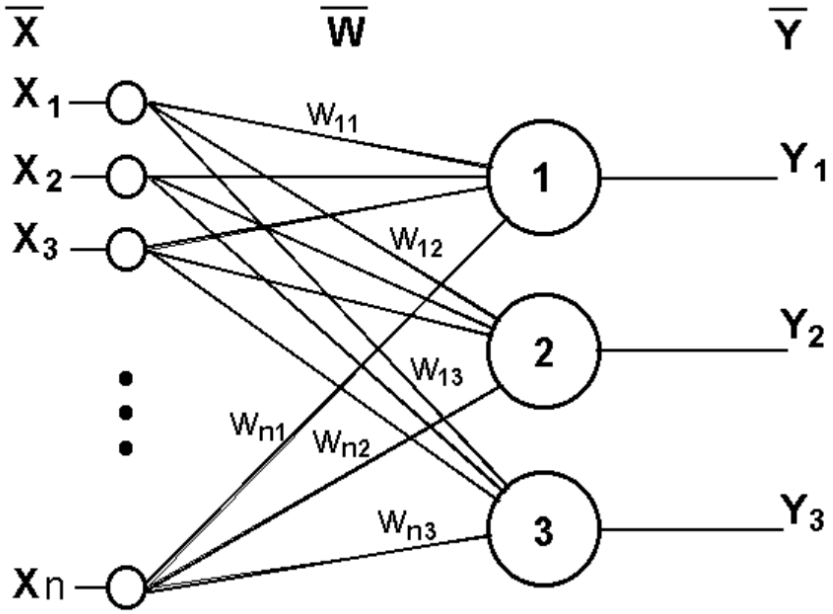
Shakl: 11.2.2. Tarmoqni o'rganish jarayoni. Qayta o`qitish hodisasi

Birinchi bosqichda mashg'ulotlar va test to'plamlari uchun prognoz xatolar bir xil bo'ladi. Keyingi bosqichlarda ikkala xatoning qiymatlari pasayadi, ammo 70-bosqichdan boshlab testlar to'plamidagi xato o'sishni boshlaydi, ya'ni. tarmoqni qayta tayyorlash jarayoni boshlanadi. Sinov to'plamidagi prognoz - bu tuzilgan modelning ishlash ko'rsatkichi. Sinov to'plamidagi xato, agar sinov to'plami hozirgi momentga iloji boricha yaqinroq bo'lsa, bashorat qilish xatosi bo'lishi mumkin.

4. Neyron tarmoq modellari

Neyron tarmoqlarining eng oddiy modellarini ko'rib chiqamiz: bir qavatli va ko'p qavatli pertseptron.

Pertseptron. Rozenblattning seminal ishida juda ko'p miqdordagi pertseptron modellari ko'rib chiqilgan. Eng oddiy neyron tarmoq modeli - bu bir qavatli perceptron. Bir qavatli pertseptron (Rozenblattning pertseptroni) - bu barcha neyronlarning qattiq aktivlashtirish funktsiyasiga ega bo'lgan bir qavatli asab tarmog'i. Bir qavatli pertseptron oddiy o'rganish algoritmiga ega va faqat eng oddiy masalalarni echishga qodir. Ushbu model 1960-yillarning boshlarida katta qiziqish uyg'otdi va sun'iy neyron tarmoqlarini rivojlanishiga turtki bo'ldi. Bunday asab tarmog'ining klassik namunasi - bitta qatlamli uch asabli pertseptron - shakl. 11.3.1.

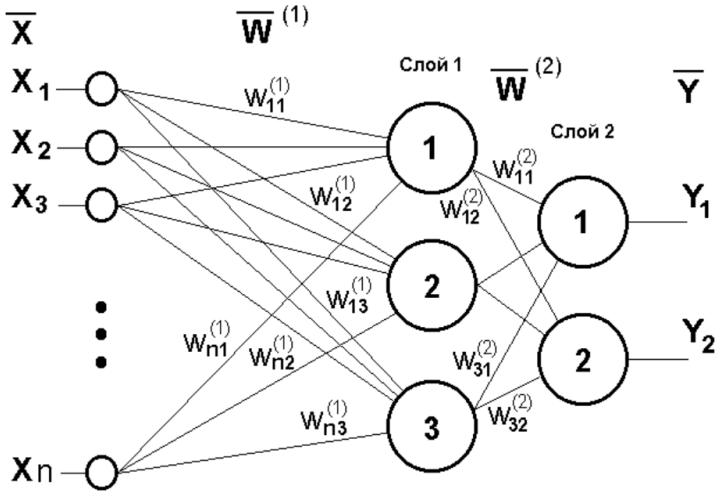


Shakl: 11.3.1. Bir qavatli uchta neyronli perseptron

Rasmda ko'rsatilgan tarmoq 3 ta neyronga sinaps orqali o'tadigan signallarni qabul qiladigan n ta kirishga ega. Ushbu uchta neyron ushbu tarmoqning yagona qatlamini tashkil qiladi va uchta chiqish signalini hosil qiladi.

Ko'p qavatli perceptron (MLP) - bu signalning to'g'ridan-to'g'ri tarqalishining neyron tarmog'i (teskari aloqasiz), unda kirish signali ketma-ket bir necha qatlamlardan o'tib, chiqishga aylanadi.

Ushbu qatlamlarning birinchisi kirish, ikkinchisi chiqish deb nomlanadi. Ushbu qatlamlar degeneratsiya deb ataladigan neyronlarni o'z ichiga oladi va ba'zida qatlamlar sonida hisobga olinmaydi. Kirish va chiqish qatlamlaridan tashqari ko'p qavatli pertseptronda bir yoki bir nechta oraliq qatlamlar mavjud bo'lib, ular yashirin qatlamlar deb ataladi. Ushbu pertseptron modeli kamida bitta yashirin qatlamga ega bo'lishi kerak. Bir nechta bunday qatlamlarning mavjudligi faqat chiziqli aktivizatsiya funktsiyalaridan foydalangan holda oqlanadi. Ikki qavatli pertseptronning namunasi shakl. 11.3.2.



Shakl: 11.3.2. Ikki qavatli pertseptron

Rasmda ko'rsatilgan tarmoq n kirishga ega. Ular sinapslar bo'ylab birinchi qavatni tashkil etuvchi 3 neyronga o'tadigan signallarni qabul qilishadi. Birinchi qavatning chiqish signallari ikkinchi qavatdagi ikkita neyronga uzatiladi. Ikkinchisi, o'z navbatida, ikkita chiqish signalini beradi.

Backpropagation (backprop) - bu xato funktsiyasi gradyanini hisoblashga asoslangan ko'p qavatli perseptronli o'rganish algoritmi. O'qitish jarayonida neyron tarmog'ining har bir qatlami neyronlarining og'irliklari oldingi qatlamdan olingan signallarni va oxirgi qatlamdan teskari yo'nalishda rekursiv ravishda hisoblangan har bir qatlamning qoldig'ini hisobga olgan holda tuzatiladi. birinchi. Boshqa neyron tarmoq modellari keyingi ma'ruzada muhokama qilinadi.

Neyron tarmoq dasturlari. Neyron tarmog'ining ishlashini simulyatsiya qiladigan dastur neyrosimulyator yoki neyropakaket deb ataladi. Ko'pgina neyro paketlar quyidagi harakatlar ketma-ketligini o'z ichiga oladi:

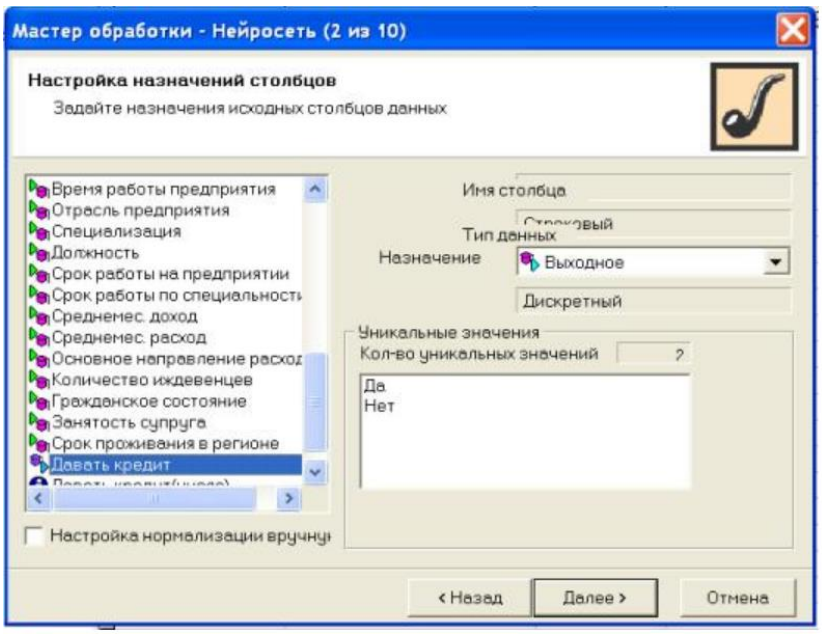
- Tarmoqni yaratish (foydalanuvchi parametrlarini tanlash yoki standart parametrlarni tasdiqlash).
- Tarmoq bo'yicha trening.
- Foydalanuvchiga qaror berish.

Neyron paketlarning xilma-xilligi juda ko'p, neyron tarmoqlardan foydalanish qobiliyati ham deyarli ma'lum bo'lgan barcha statistik paketlarga kiritilgan. Ixtisoslashtirilgan neyropaketlar qatoriga quyidagilar kiradi: BrainMaker, NeuroOffice, NeuroPro va boshqalar Neyropakaketlarni taqqoslash mezonlari:

foydalanish qulayligi, taqdim etilgan ma'lumotlarning ravshanligi, turli xil tuzilmalardan foydalanish qobiliyati, ishlash tezligi, hujjatlar mavjudligi. Tanlov foydalanuvchining malakasi va talablari bilan belgilanadi.

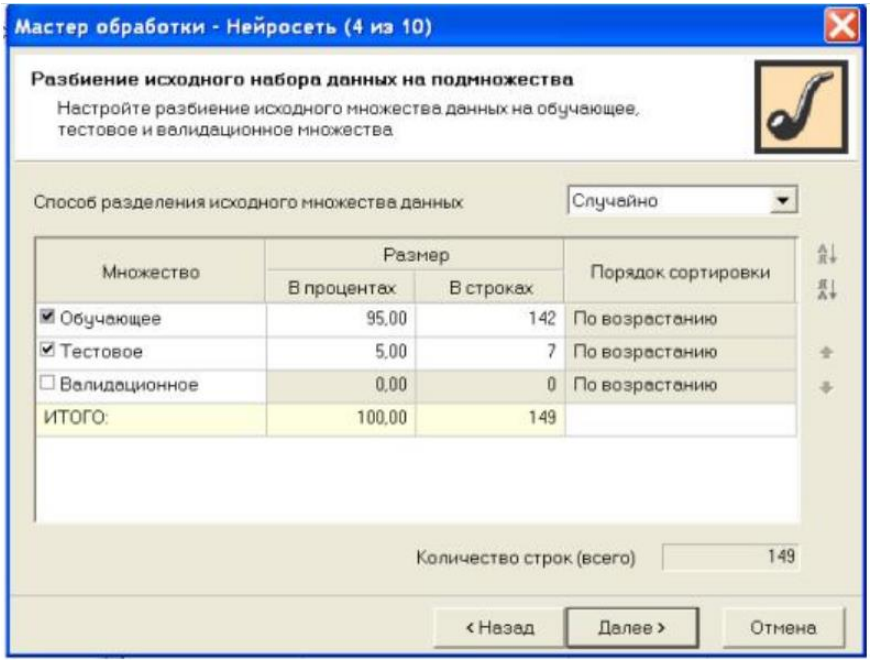
Muammoni hal qilishning misoli

Deductor (BaseGroup) analitik to'plamida "Mijozga qarz berish yoki bermaslik" muammosining echimini ko'rib chiqamiz. O'quv ma'lumotlar bazasi - bu mijozlar to'g'risidagi ma'lumotlarni o'z ichiga olgan ma'lumotlar bazasi, xususan: kredit miqdori, qarz muddati, kredit maqsadi, yoshi, jinsi, ma'lumoti, xususiy mulk, kvartira, kvartira maydoni. Ushbu ma'lumotlarga asoslanib, kredit olishni istagan Mijozning kreditni qaytarish xavfi guruhiga kiradimi-yo'qligiga javob beradigan modelni yaratish kerak, ya'ni. foydalanuvchi "Qarz berishim kerakmi?" degan savolga javob olishi kerak. Vazifa tasniflash vazifalari guruhiga kiradi, ya'ni. o'qituvchi bilan o'rganish. Tahlil uchun ma'lumotlar credit.txt faylida joylashgan. Fayldan ma'lumotlarni import qilish ustasi yordamida import qilamiz. Biz ishlov berish ustasini ishga tushiramiz  va ma'lumotlarni qayta ishlash usulini - neyron tarmoqni tanlaymiz. Manba ma'lumotlari ustunlari yo'nalishini belgilang. Bizning vazifamizdagi chiqish ustuni "Kredit bering", qolganlarning hammasi kirishdir. Ushbu qadam shakl. 11.3.3.



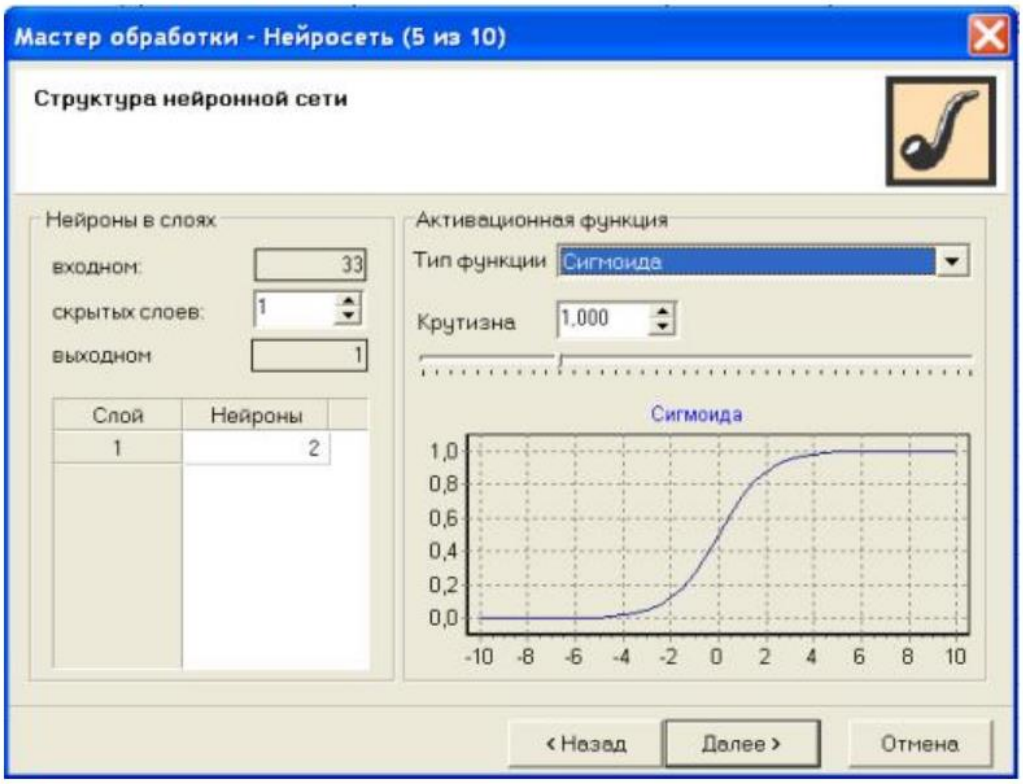
Shakl: 11.3.3. Ustunlarni tayinlash bosqichini sozlash

Keyingi bosqichda dastur dastlabki ma'lumotlar to'plamini o'qitish va sinovga ajratishni taklif qiladi. Asl ma'lumotlar to'plami uchun standart bo'linish usuli "Tasodifiy". Ushbu qadam shakl. 11.3.4.



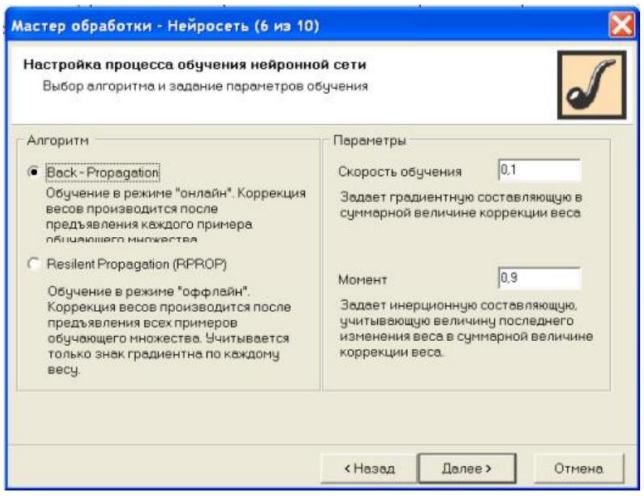
Shakl: 11.3.4. "Asl ma'lumotlar to'plamini pastki to'plamlarga ajratish" bosqichi

Keyingi qadam neyron tarmog'ining tuzilishini aniqlashdir, ya'ni, kirish qatlamidagi neyronlar sonini - 33 (kirish o'zgaruvchilar soni), yashirin qatlamda - 1, chiqish qatlamida - 1 (chiqish o'zgaruvchilar soni) ni ko'rsating. Aktivizatsiya funktsiyasi Sigmoid bo'lib, uning qiyaligi biriga teng. Ushbu qadam shakl. 11.3.5.



Shakl: 11.3.5. "Neyron tarmoq tuzilishi" bosqichi

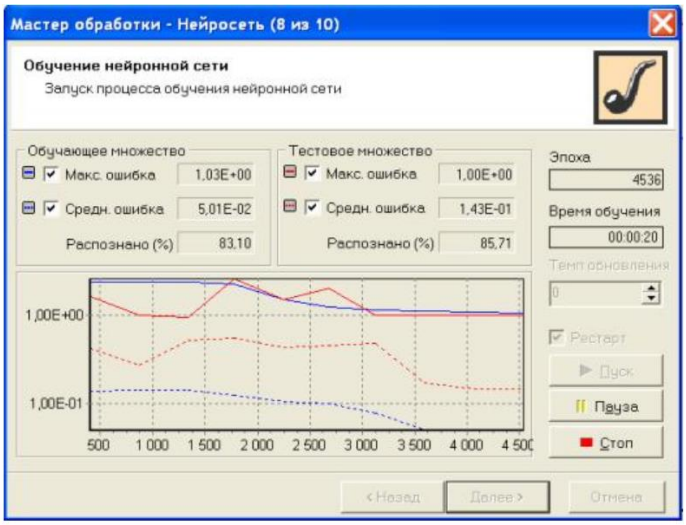
Keyinchalik, biz neyron tarmog'ini o'qitish uchun algoritm va parametrlarni tanlaymiz. Ushbu qadam "Neyron tarmog'ini o'qitish jarayonini sozlash" deb nomlangan, bu rasmda ko'rsatilgan. 11.3.6.



Shakl: 11.3.6. "Neyron tarmog'ini o'qitish jarayonini sozlash" bosqichi

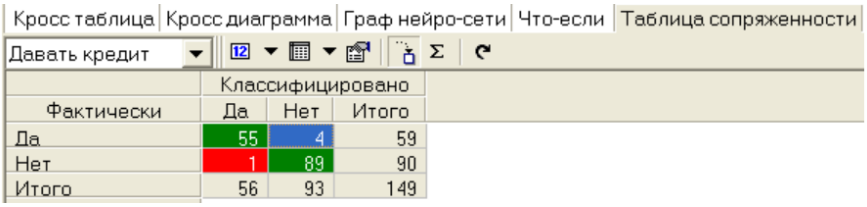
Keyingi bosqichda biz mashg'ulotni to'xtatish shartlarini o'rnatdik. Agar xato 0,005 dan kam bo'lsa, biz tan olingan misolni ko'rib chiqamiz va 10000 davriga etganimizda mashg'ulotni to'xtatish shartini ko'rsatamiz, keyingi bosqichda biz mashg'ulot jarayonini boshlaymiz va xato qiymati va foizning o'zgarishini kuzatamiz. o'quv va test to'plamlarida tan olingan misollar. Bizning holatlarimizda

4536-sonli davrda 83,10% namunalar o'quv majmuasida, 85,71% namunalar test to'plamida tan olinganligini ko'ramiz. Ushbu jarayonning bir qismi shakl. 11.3.7.



Shakl: 11.3.7. "Neyron tarmoqlari bo'yicha trening" bosqichi

O'quv jarayoni tugagandan so'ng, olingan natijalarni talqin qilish uchun biz taklif etilayotganlar ro'yxatidan vizualizatorlarni tanlash imkoniyatiga egamiz. Keling, quyidagilarni tanlaymiz: favqulodda vaziyatlar jadvali, neyron tarmoqlari grafigi, agar bo'lsa tahlil qilish va ulardan olingan ma'lumotlarni tahlil qilish uchun foydalaning. Shakl. 11.3.8. kutilmagan holatlar jadvalini ko'rsatadi. Uning diagonalida to'g'ri tan olingan misollar mavjud, ya'ni. Sizga qarz berishingiz mumkin bo'lgan 55 ta mijoz va qarz bermasligi kerak bo'lgan 89 ta mijoz. Qolgan kataklar boshqa sinfga (1 va 4) ajratilgan mijozlarni o'z ichiga oladi. Deyarli barcha misollar to'g'ri tasniflangan deb hisoblash mumkin - 96,64%.



Shakl: 11.3.8. Favqulodda vaziyatlar jadvali

Vizualizator sizga tajriba o'tkazishga imkon beradi. Potentsial qarz oluvchi to'g'risidagi ma'lumotlar tegishli maydonlarga kiritilishi kerak va qurilgan model "Kredit berish" maydonining qiymatini hisoblab chiqadi - "Ha" yoki "Yo'q", ya'ni, muammoni hal qiladi.

Matlab to'plami. MATLAB (MathWorks) to'plami, shuningdek, foydalanuvchilarga neyron tarmoqlari bilan ishlash imkoniyatini beradi. MATLAB standart etkazib berishga kiritilgan "Neyron Tarmoq asboblar qutisi" barcha turdagi neyron tarmoqlari bilan ishlash uchun keng imkoniyatlar yaratadi. MATLAB paketining afzalligi shundaki, undan foydalanishda foydalanuvchi neyrosimulyatorga qattiq joylashtirilgan neyron tarmoqlari modellari va ularning parametrlari bilan cheklanib qolmaydi, balki uni echish uchun maqbul deb hisoblagan tarmoqni mustaqil ravishda loyihalashtirishga qodir. muammo. Matlab paketida neyron tarmoqlarini qurish misolini ko'rib chiqamiz. 15 ta mustaqil o'zgaruvchi bo'lsin - firma faoliyati ko'rsatkichlari va bitta bog'liq o'zgaruvchi - sotuvlar. O'tgan yil uchun ma'lumotlar bazamiz bor. Bir oy davomida sotish hajmining haftalik prognozini tuzish kerak. Muammoni hal qilish uchun uch qavatli backpropagation tarmog'idan foydalanish taklif etiladi. Kirish qavatidagi 15 ta neyronni (kiritilgan o'zgaruvchilar soni bo'yicha), ikkinchi qavatdagi 8 ta neyronni va chiqish qavatidagi 1 ta neyronni (chiqadigan o'zgaruvchilar soni bo'yicha) o'z ichiga olgan tarmoq hosil qilaylik.

Har bir qatlam uchun biz uzatish funktsiyasini tanlaymiz: birinchi qavat tansig, ikkinchisi loggsig, uchinchisi purelin. Matlabda bunday asab tarmog'ining sintaksisi quyidagicha:

```
Net=netff(PR, [S1,S2, : , Sn],{TF1,TF2, : , TFn},btf,
blf, pf),
```

bu erda PR - R kirish vektorlari uchun minimal va maksimal qiymatlar massivi;

Si - i-qavatdagi neyronlarning soni;

TFi - bu i qatlamini faollashtirish funktsiyasi;

btf - backpropagation usulini amalga oshiradigan o'quv funktsiyasi;

blf - bu backpropagation usulini amalga oshiradigan konfiguratsiya funktsiyasi;

pf - bu kadrlar tayyorlash sifati mezonidir.

Har qanday farqlanadigan funktsiya aktivizatsiya vazifasini bajarishi mumkin, masalan, tansig, loggsig, purelin.

```
Net=netff(minmax (P), [n,m, l],{ tansig, logsig,
purelin },trainpr),
```

bu erda P - kirish vektorlari to'plami;

n - NS kirishlar soni;

m - yashirin qatlamdagi neyronlarning soni;

I - NS chiqishlari soni.

Shuningdek, xato qiymatini hisoblash usulini o'rnatish kerak. Masalan, eng kichik kvadratchalar usuli tanlangan bo'lsa, unda bu funktsiya quyidagicha ko'rinadi: Net.performFcn = 'SSE'. Maksimal davrlar sonini 10000 ga o'rnatish uchun quyidagi funktsiyadan foydalaning: net.trainParam.epochs = 10,000. Siz o'quv jarayonini shu tarzda boshlashingiz mumkin:

```
[net,tr]=train(net,P,T);
```

Tarmoqni o'qitishni tugatgandan so'ng, u faylda saqlanishi mumkin, masalan, nn1.mat nomi bilan.

Buning uchun quyidagi buyruqni bajaring:

```
save nn1 net;
```

Shunday qilib, paketda har qanday murakkablikdagi tarmoqni loyihalash mumkin va neyrosimulyatorlar tomonidan qo'yiladigan cheklovlarga bog'lanishning hojati yo'q. Shu bilan birga, Matlab paketidagi neyron tarmoqlari bilan ishlash uchun atrof muhitning o'zi va asab tizimining asboblar qutisi funktsiyalarining ko'pini o'rganish kerak. Neyron tarmog'ining asboblar qutisidagi neyron tarmoqlari dizaynini batafsil o'rganish uchun quyidagilarni tavsiya etish mumkin.

Nazorat savollari:

1. Neyron tarmog`i nima?
2. Neyron tarmoq dasturlariga misollar keltiring.
3. Qatlamli tarmoqning qanday turlarini bilasiz?
4. Qanday faollashtirish funksiyalarini bilasiz?
5. Neyron tarmog`lar asosan qaysi sohalarda keng qo`llaniladi?

12-MAVZU

OLAP VA BOSHQA MA`LUMOT SAQLAGICHLAR. OLAP VA DATA MINING INTEGRATSIYASI

Reja:

1. **Ma`lumot saqlagichalar**
2. **OLAP tizimlari**
3. **OLAP va Data Mining integratsiyasi**

**Mashg`ulot maqsadi:** *Mashg`ulotda axborot tizimlari, ularning turlari va tarkibiy qismlari muhokama qilinadi. OLAP-texnologiyasining asosiy g'oyalari, OLAP-server arxitekturasi, Data Mining va OLAP integratsiyasi. Ma'lumotlar omborlari texnologiyasi va ulardan foydalanishning afzalliklari, xususan, Data Mining jarayoni uchun tavsiflangan.*

Tayanch iboralar: *dasturiy ta'minot, Data Mining , axborot piramidasi, qaror qabul qilish, qarorlarni qo'llab-quvvatlash tizimi, DSS, ta'rif, tajriba, avtomatlashtirilgan tizim, qaror qabul qiluvchi, yarim tuzilgan, tuzilmagan, tuzilgan, tuzilmagan vazifa, tuzilgan vazifa, yarim tuzilgan vazifa, ma'lumotlar ombori , OLAP, tahlil, quvvat, ma'lumotlarga asoslangan, DSS, EIS, ijro, axborot tizimi, WIS, interfeys, qarorlarni qo'llab-quvvatlash tizimi, domen, vaqtinchalik, qo'llab-quvvatlash, bo'linish, analitik ishlov berish, taqdimot, ko'p o'lchovli, kontseptual ko'rinish, foydalanuvchi, operatsiyalar, MOLAP, ROLAP, HOLAP, gibrid, gibrid arxitektura, dastur, ma'lumotlarni saqlash, server, ma'lumotlarni o'zgartirish, kesh, integratsiya, kirish, qidirish, ehtimollik, kombinatsiyalar, HAN, axborot tizimlari, ma'lumotlar bazalari, ma'lumotlar, ob’yekt, atribut, yulduz.*

1. Ma`lumot saqlagichalar

Avvalgi mashg`ulotlarning birida biz ma'lumot piramidasini ko'rib chiqdik, bu harakat davomida ma'lumotlar hajmidan echimlarga qadar bilim hajmi biznes qiymatiga aylanadi. Ushbu ma'lumot piramidasini yuqoriga ko'tarish bilan bog'liq bo'lgan Data Mining jarayoni qarorlarni qabul qilish jarayoni bilan uzviy bog'liqdir,

uni qarorlarni qo'llab-quvvatlash tizimlarining (QQQT) ajralmas qismi deb hisoblash mumkin.

Shunday qilib, Data Mining qarorlarni qo'llab-quvvatlash jarayoni sifatida qaralishi mumkin, shu bilan birga to'plangan ma'lumotlar avtomatik ravishda bilim sifatida tavsiflanishi mumkin bo'lgan ma'lumotlarga umumlashtiriladi. Qarorlar va qarorlar kontseptsiyasi bilan biz kursning dastlabki ma'ruzalaridan birida qisqacha tanishib chiqdik. QQQT o'tgan asrning 70-yillari boshlarida boshqaruvning axborot tizimlari va ma'lumotlar bazasini boshqarish tizimlarining rivojlanishi natijasida paydo bo'ldi. Hozirgi vaqtda inson faoliyatining turli sohalarida ishlab chiqilgan va amalga oshirilgan juda ko'p sonli QQQTlar mavjud. Ularning rivojlanish tezligi doimiy ravishda oshib bormoqda. Biroq, bugungi kunda, ushbu tizimlarning keng tarqalishiga qaramay, ushbu atamaning umumiy qabul qilingan ta'rifi hali topilmadi. Shuni ta'kidlash kerakki, QQQT butun dunyoda keng qo'llanilgan bo'lsa-da, MDHda ushbu turdagi tizimlarga hali etarlicha e'tibor berilmagan. Keling, qarorlarni qo'llab-quvvatlash tizimi nima ekanligini ko'rib chiqaylik. Yuqorida ta'kidlab o'tilganidek, ushbu masala, shuningdek QQQT sinfiga har xil turdagi tizimlarni kiritish masalasi munozarali; bu boradagi fikrlar ko'pincha hatto bir-biriga zid keladi. QQQT ning ba'zi ta'riflari. QQQT boshqaruv muammolarini hal qilish tajribasini o'z ichiga olgan va oqilona qarorlarni ishlab chiqish jarayonida mutaxassislar guruhining ishtirokini ta'minlaydigan tadqiqot, ekspert va aqlli tizimlar uchun tegishli axborot ta'minoti bilan o'zaro bog'liq modellar to'plamiga asoslanadi. Qarorlarni qo'llab-quvvatlash tizimi bu qarorlar qoidalari va ma'lumotlar bazalari bilan mos keladigan modellardan foydalanadigan interaktiv avtomatlashtirilgan tizim, shuningdek, kompyuterning interaktiv simulyatsiyasi jarayonidir. QQQT "qarorlarni hisoblashda" vosita bo'lib, u "qaror qabul qilishda (bundan buyon matnda - QQ) qaror qabul qilishda yordam beradigan ma'lumotlarni qayta ishlashning bir qator protseduralari va qarorlaridan foydalanishga" asoslangan. QQQT "qaror qabul qiluvchilarga ma'lumotlar va modellardan tuzilmaviy muammolarni hal qilishda foydalanishda yordam beradigan interaktiv avtomatlashtirilgan tizimlar" dir. QQQT - "butun qaror qabul qilish jarayonini to'liq amalga oshiradigan avtomatik tizimlarga ega bo'lish imkonsiz yoki

nomaqbul bo'lgan holatlarda qaror qabul qilishda turli xil faoliyatni qo'llab-quvvatlash uchun ishlatiladigan kompyuter axborot tizimi." QQQT qaror qabul qiluvchining o'rnini bosmaydi, qaror qabul qilish jarayonini avtomatlashtiradi, lekin unga vazifani hal qilishda yordam beradi. Shuni ta'kidlash kerakki, QQQTning birinchi ta'riflaridan boshlab, ularning yordami bilan hal qilingan vazifalar doirasi yarim tuzilgan va tuzilmasiz vazifalar bilan cheklangan edi. QQQT ni quyidagicha ta'riflaylik: QQQT - bu inson faoliyatining har xil turdagi yarim tuzilgan va tuzilmaviy muammolarida qaror qabul qilishni qo'llab-quvvatlash uchun mo'ljallangan interfaol kompyuter tizimi.

Ushbu ta'rifning muhim tushunchalari:

- kompyuter interaktiv (ya'ni, qarorni qo'llab-quvvatlash tizimining qaror qabul qiluvchisi tomonidan bevosita foydalanishni talab qilmaydigan);
- qarorni qo'llab-quvvatlash (qaror shaxs tomonidan qabul qilinadi);
- yarim tuzilgan va tuzilmagan muammolar (bu menejerlar hal qiladigan muammolar).

Muammolarni yarim tuzilishga, tuzilmaga va tuzilmalarga ajratish nima ekanligini ko'rib chiqamiz.

Tuzilmaviy vazifalar qaror qabul qiluvchining qarorlari asosida faqat sifatli tavsifga ega, vazifaning asosiy xususiyatlari o'rtasidagi miqdoriy munosabatlar ma'lum emas.

Tuzilgan vazifalar miqdoriy jihatdan aniqlanishi mumkin bo'lgan muhim bog'liqliklar bilan tavsiflanadi.

Zaif tuzilgan vazifalar oraliq pozitsiyani egallaydi va "miqdoriy va sifat bog'liqliklarini birlashtiradi va vazifaning kam ma'lum va noaniq tomonlari ustunlik qiladi".

QQQT klassik tuzilishining asosini tashkil etuvchi uchta komponent mavjud bo'lib, uni boshqa turdagi axborot tizimlaridan ajratib turadi: foydalanuvchi interfeysi quyi tizimi, ma'lumotlar bazasini boshqarish quyi tizimi va model bazasini boshqarish quyi tizimi. Agar QQQT -ni funktsional tomondan ko'rib chiqib, quyidagi komponentlarni ajratib ko'rsatish mumkin:

- ma'lumotlar ombori serveri;
- OLAP uskunalar to'plami;
- Data Mining bo'yicha qo'llanma.

QQQT ning ushbu tarkibiy qismlari quyidagi asosiy masalalarni hal qiladi: ma'lumotlarni kontseptsiya darajasida to'plash va modellashtirish masalasi, bir nechta mustaqil manbalardan ma'lumotlarni samarali yuklash va ma'lumotlarni tahlil qilish masalasi. Aytishimiz mumkinki, bugungi kunda onlayn analitik ishlov berish (OLAP tizimlari) ko'p o'lchovli ma'lumotlarga kirishni ta'minlash bilan cheklangan. Ma'lumotlarni tanlab olish texnologiyasi QQQT uchun eng katta qiziqish uyg'otadi, chunki u ma'lumotlarning eng chuqur va keng qamrovli tahlilini o'tkazish va shu sababli eng muvozanatli va asosli qarorlarni qabul qilish uchun ishlatilishi mumkin.

QQQT tasnifi. QQQT tasniflari masalasi hozirgi kunda dolzarb bo'lib, yangi taksonomiyalarni ishlab chiqish davom etmoqda. Keling, ulardan ikkitasini ko'rib chiqaylik. Quyida DSSning ayrim xususiyatlarining o'xshashligiga qarab tasnifi keltirilgan (D.J. Power, 2000). Guruhlarning batafsil tavsifi bilan tanishishingiz mumkin.

- QQQT, ma'lumotlarga yo'naltirilgan (Data-driven DSS, Data-oriented DSS);
- QQQT, modelga asoslangan (Model-driven DSS);
- DSS, bilimga yo'naltirilgan (Knowledge-driven DSS);
- QQQT, hujjatlarga asoslangan (Document-driven DSS);
- Aloqa va QQQT guruhiga yo'naltirilgan DSS (Communications-Driven? Group DSS);
- Tashkilotlararo va uyushgan QQQT (Inter-Organizational или Intra-Organizational DSS);
- Maxsus funktsional QQQT yoki umumiy maqsadli DSS (Function-Specific или General Purpose DSS);
- Internetga asoslangan QQQT (Web-Based DSS).
- QQQTlar ishlaydigan ma'lumotlarga qarab ikkita asosiy DSS mavjud: EIS va DSS.

EIS (Execution Information System) - boshqaruv axborot tizimi, EIS.

Ushbu turdagi DSS amaldagi vaziyatga zudlik bilan javob berish uchun mo'ljallangan. Ularning aksariyati tayyor bo'lmagan foydalanuvchiga qaratilgan, shuning uchun ular soddalashtirilgan interfeysga, taklif etilayotgan qobiliyatlarning asosiy to'plamiga, axborotni taqdim etishning aniq shakllariga va echilishi kerak bo'lgan vazifalar ro'yxatiga ega. Bunday tizimlar odatdagi so'rovlarga asoslanadi, ular nisbatan kam; bunday so'rovlar natijasida olingan hisobotlar eng qulay shaklda taqdim etiladi.

DSS (Qarorlarni qo'llab-quvvatlash tizimi). Ushbu turdagi tizimlarga ko'p funktsiyali ma'lumotlarni tahlil qilish va tadqiqot tizimlari kiradi. Ular qaror qabul qilishda ishlatilishi mumkin bo'lgan ma'lumotlarni chuqur qayta ishlashni o'z ichiga oladi.

Ushbu turdagi tizimlar, EISdan farqli o'laroq, mavzular bo'yicha ham bilimga, ham zamonaviy kompyuter texnologiyalaridan foydalanish qobiliyatiga ega foydalanuvchilar uchun mo'ljallangan. Ushbu tizimlar sun'iy intellektning xususiyatlari bilan tavsiflanadi, chunki topshiriq bo'yicha dastlabki ma'lumotlarni aniq xulosalar qilish imkoniyati mavjud. Agar ma'lumotlarni umumlashtirish va tahlil qilish va ularni qayta ishlash uchun asoslar mavjud bo'lsa, bunday tizimlarni yaratish mantiqan to'g'ri keladi.

Yaqinda DSS-ga faqat ikkinchi turga tegishli bo'lgan, ya'ni. DSS. Ushbu turdagi tizimlar ba'zan dinamik tizimlar deb ataladi, ya'ni. ular kutilmagan (maxsus) so'rovlarni ko'rib chiqishga yo'naltirilgan bo'lishi kerak. To'plangan ma'lumotlarga asoslangan qarorlarni qo'llab-quvvatlash uchta asosiy yo'nalishda amalga oshirilishi mumkin.

1. Batafsil ma'lumotlar maydoni (OLTP tizimlari). Ushbu tizimlarning ko'pchiligining maqsadi - ma'lumotlarni qidirish, bular axborot qidirish tizimlari deb ataladi. Ular ma'lumotlarni qayta ishlash tizimlarida qo'shimcha sifatida yoki ma'lumotlarni saqlash sifatida ishlatilishi mumkin.
2. Birlashtirilgan ko'rsatkichlar sohasi (OLAP tizimlari). OLAP tizimlarining vazifalari umumlashtirish, yig'ish, ma'lumotni giperkubik taqdim etish va ko'p

o'zgaruvchan tahlildir. Ular ko'p o'lchovli DBMS yoki ma'lumotlarning dastlabki yig'ilishi bilan relyatsion ma'lumotlar bazalari bo'lishi mumkin.

3. Qonuniyatlar sohasi (Data Mining).

Tizimlarning EIS va DSS ga bunday bo'linishi DSS turlaridan birini amalga oshirishni anglatmaydi. Tizimlarning har biri ma'lum bir toifadagi foydalanuvchilarga o'z funktsiyalarini taqdim etganda, ular parallel ravishda mavjud bo'lishi mumkin. Qarorni qo'llab-quvvatlashning umumiy sxemasi quyidagilarni o'z ichiga oladi:

- qaror qabul qiluvchilarga boshqariladigan tizimning holatini va unga ta'sirini baholashda yordam berish; qaror qabul qiluvchilarning afzalliklarini aniqlash;
- mumkin bo'lgan yechimlarni yaratish;
- qaror qabul qiluvchining afzalliklari asosida mumkin bo'lgan alternativalarni baholash;
- qabul qilingan qarorlarning natijalarini tahlil qilish va qaror qabul qiluvchi nuqtai nazaridan eng yaxshisini tanlash.

2. OLAP tizimlari

OLAP yoki On-layn analitik ishlov berish kontseptsiyasi ko'p o'lchovli kontseptual ko'rinishga asoslangan. OLAP atamasi 1993 yilda E. F. Kodd tomonidan kiritilgan. Ushbu tizimning asosiy g'oyasi foydalanuvchilar tomonidan kirish mumkin bo'lgan ko'p o'lchovli jadvallarni yaratishdir. Ushbu ko'p o'lchovli jadvallar yoki ko'p o'lchovli kublar deb ataladigan narsalar manba va yig'ilgan ma'lumotlardan tuzilgan. Ko'p o'lchovli jadvallar uchun ham manba, ham yig'ilgan ma'lumotlar relyatsion va ko'p o'lchovli ma'lumotlar bazalarida saqlanishi mumkin. OLAP-tizim bilan o'zaro aloqada bo'lib, foydalanuvchi ma'lumotni moslashuvchan ko'rishni amalga oshirishi, har xil ma'lumotlar bo'laklarini olishi, detallashtirish, bir uchidan ikkinchi uchigacha taqsimlash, vaqt o'tishi bilan taqqoslash bo'yicha analitik operatsiyalarni bajarishi mumkin. OLAP-tizim bilan barcha ishlar predmet sohasi nuqtai nazaridan sodir bo'ladi.

OLAP mahsulotlari. Bugungi kunda bozorda juda ko'p turli xil OLAP tizimlari mavjud. Ushbu turdagi mahsulotlarning bir nechta tasniflari ishlab chiqilgan:

masalan, ma'lumotlarni saqlash usuli bo'yicha, OLAP mashinasining joylashuvi bo'yicha, foydalanishga tayyorlik darajasiga ko'ra tasniflash. Yuqoridagi tasniflardan birinchisini ko'rib chiqing. OLAP tizimlarida ma'lumotlarni saqlashning uchta usuli yoki uchta OLAP server arxitekturasi mavjud:

- **MOLAP (ko'p o'lchovli OLAP);**
- **ROLAP (Relational OLAP);**
- **HOLAP (gibrid OLAP).**

Shunday qilib, ushbu tasnifga ko'ra, OLAP mahsulotlarini uchta sinf tizimlari ko'rsatish mumkin.

- MOLAP uchun manba va ko'p o'lchovli ma'lumotlar ko'p o'lchovli ma'lumotlar bazasida yoki ko'p o'lchovli mahalliy kubda saqlanadi. Ushbu saqlash usuli OLAP operatsiyalarini bajarilishining yuqori tezligini ta'minlaydi. Ammo bu holda ko'p o'lchovli baza ko'pincha keraksiz bo'ladi. Uning asosida qurilgan kub o'lchovlar soniga juda bog'liq bo'ladi. O'lchamlarning soni oshgani sayin, kub hajmi shiddat bilan o'sib boradi. Ba'zan bu ma'lumotlar hajmining "portlovchi o'sishiga" olib kelishi mumkin, natijada foydalanuvchi so'rovlari falajlanadi.
- ROLAP mahsulotlarida manba ma'lumotlar relyatsion ma'lumotlar bazalarida yoki fayl serveridagi tekis mahalliy jadvallarda saqlanadi. Umumiy ma'lumotlar bir xil ma'lumotlar bazasidagi xizmat jadvallariga joylashtirilishi mumkin. Ma'lumotlarni relyatsion ma'lumotlar bazasidan ko'p o'lchovli kublarga o'tkazish OLAP vositasi talabiga binoan amalga oshiriladi. Shu bilan birga, kubni qurish tezligi ma'lumotlar manbai turiga juda bog'liq bo'ladi va shuning uchun tizimning javob vaqti ba'zan qabul qilinishi mumkin bo'lmagan darajada uzoqlashadi.
- Gibrid arxitekturadan foydalanish holatida, ya'ni. HOLAP mahsulotlarida asl ma'lumotlar relyatsion ma'lumotlar bazasida qoladi va agregatlar ko'p o'lchovli ma'lumotlar bazasiga joylashtiriladi. OLAP kubini yaratish OLAP vositasi talabiga binoan relyatsion va ko'p o'lchovli ma'lumotlarga asoslanib amalga oshiriladi. Ushbu yondashuv portlovchi ma'lumotlarning o'sishini

oldini oladi. Shu bilan birga, siz mijoz so'rovlarining maqbul bajarilish vaqtiga erishishingiz mumkin.

Keyingi tasnif OLAP mashinasining joylashishiga asoslangan. Shu asosda OLAP mahsulotlari OLAP serverlari va OLAP mijozlariga bo'linadi. Server OLAP vositalarida yig'ilgan ma'lumotlarni hisoblash va saqlash alohida jarayon - server tomonidan amalga oshiriladi. Mijozlar dasturi faqat serverda saqlanadigan ko'p o'lchovli kublar bo'yicha so'rovlar natijalarini oladi. Ba'zi OLAP serverlari ma'lumotlarni faqat relyatsion ma'lumotlar bazalarida, boshqalari faqat ko'p o'lchovli ma'lumotlar bazalarida saqlashni qo'llab-quvvatlaydi. Ko'pgina zamonaviy OLAP serverlari barcha uchta saqlash usullarini qo'llab-quvvatlaydi: MOLAP, ROLAP va HOLAP. Bugungi kunda eng keng tarqalgan server echimlaridan biri bu Microsoft OLAP-serverdir. OLAP mijozi boshqa tuzilishga ega. O'lchamli kubni yaratish va OLAP hisob-kitoblari mijoz kompyuterining xotirasida amalga oshiriladi. OLAP-server yordamida qayta ishlangan ko'p o'lchovli ma'lumotlarning jismoniy saqlanishi tashkil etilishi mumkin [81], bu foydalanuvchi so'rovlariga tezkor javob berishga imkon beradi. Bundan tashqari, u ma'lumotlarning relyatsion va boshqa ma'lumotlar bazalaridan real vaqt rejimida ko'p o'lchovli tuzilmalarga aylanishini ta'minlaydi. Relatsion va ko'p o'lchovli vositalar qanday ishlaydi? OLAP mahsulotlari relyatsion tizimlar bilan integratsiya qilish orqali mavjud korporativ infratuzilma bilan birlashtirilmoqda. Ma'lumotlar bazasi ma'murlari aloqador ma'lumotlarni ko'p o'lchovli keshga yuklashadi yoki SQL ma'lumotlariga kirish uchun keshni sozlashadi.

12.2.1-jadvalda ma'lumotlarni boshqarishning turli modellarining qiyosiy xususiyatlari ko'rsatilgan.

Xususiyatlari	**Relyatsion DBMS OLTP**	**Relatsion DBMS DSS / ma'lumotlar ombori**	**Ko'p o'lchovli OLAP ma'lumotlar bazasi**
Odatiy operatsiya	Yangilash	Отчет	Tahlil
Analitik talablar darajasi	Past	O'rtacha	Yuqori

Oynalar	O'zgarmas	Foydalanuvchi tomonidan belgilangan	Foydalanuvchi tomonidan belgilangan
Bitta operatsiya bo'yicha ma'lumotlar hajmi	Katta bo`lmagan	Kichikdan kattagacha	Katta
Ma'lumotlar qatlami	Batafsil	Batafsil va jami	Jami
Ma'lumotlarni saqlash muddatlari	Только текущие	Oldingi va hozirgi	Tarixiy, hozirgi va prognoz qilingan
Strukturaviy elementlar	Yozuvlar	Yozuvlar	Massivlar

12.1.1-jadval. Ma'lumotlarni boshqarish turli modellarining qiyosiy xususiyatlari

3. OLAP va Data Mining integratsiyasi

Ikkala texnologiyani ham qarorlarni qo'llab-quvvatlash jarayonining ajralmas qismlari deb hisoblash mumkin. Biroq, bu texnologiyalar turli yo'nalishlarda harakatlanayotganga o'xshaydi: OLAP faqat ko'p o'lchovli ma'lumotlarga kirishni ta'minlashga qaratilgan va Data Mining usullari aksariyat hollarda bir o'lchovli jadvallar va relyatsion ma'lumotlar bilan ishlaydi. OLAP va Data Mining texnologiyalarining integratsiyasi har ikkala texnologiyaning funksionalligini "boyitadi". Ushbu ikki turdagi tahlillar birlashtirilishi kerak, shunda integral texnologiya bir vaqtning o'zida ko'p o'lchovli kirish va naqshlarni qidirishni ta'minlaydi. N. Radenning so'zlariga ko'ra, "ko'plab kompaniyalar ... ma'lumotlarning yaxshi omborlarini yaratdilar, tokchalarda foydalanilmaydigan ma'lumotlarning tog'larini ideal tarzda saralashdi, bu o'z-o'zidan bozor voqealariga tez yoki etarlicha vakolatli munosabat bildirmaydi".

K. Parsaye bunday kombinatsiyani ko'rsatish uchun "OLAP Data Mining" (ko'p o'lchovli Data Mining) birikma atamasini kiritadi.

Data Miningning ko'p o'zgaruvchan vositasi har xil umumlashma darajalariga ega bo'lgan batafsil va jamlangan ma'lumotlarda naqshlarni topishi kerak. Ko'p o'lchovli ma'lumotlarni tahlil qilish maxsus hujayralar giperkubasi asosida tuzilishi kerak, ularning hujayralarida o'zboshimchalik bilan raqamli qiymatlar (hodisalar soni, savdo hajmi, yig'ilgan soliqlar miqdori) mavjud emas,

lekin mos keladigan ehtimollikni aniqlaydigan raqamlar atribut qiymatlarining kombinatsiyasi. Bunday giperkubaning proektsiyalari (individual o'lchovlarni hisobga olishdan tashqari) ham qonuniyatlarni topish uchun tekshirilishi kerak. J. Xan yanada sodda nom - "OLAP Mining" ni taklif qiladi va ikkita texnologiyani birlashtirish uchun bir nechta variantni taklif qiladi.

1. "Cubing then mining". Kon tahlilini o'tkazish qobiliyati so'rovning har qanday natijasi bo'yicha ko'p o'lchovli kontseptual tasvirga berilishi kerak, ya'ni ko'rsatkichlarning giperkubasining har qanday proektsiyasining har qanday bo'lagi ustida.
2. "Mining then cubing". Ombordan olingan ma'lumotlar singari, kon qazish natijalari ham keyingi ko'p o'zgaruvchan tahlil uchun giperkubik shaklda taqdim etilishi kerak.
3. "Cubing while mining". Ushbu moslashuvchan moslashuv usuli ko'p o'zgaruvchan tahlilning har bir bosqichi (umumlashtirish darajalari o'rtasida o'tish, giperkubaning yangi qismini olish va hk) natijasida bir xil turdagi aqlli ishlov berish mexanizmlarini avtomatik ravishda faollashtirishga imkon beradi.

Bugungi kunga kelib, bir nechta ishlab chiqaruvchilar ko'p o'lchovli ma'lumotlar uchun Data Mining dasturini amalga oshirmoqdalar. Bundan tashqari, ba'zi bir Data Mining usullari, masalan, eng yaqin qo'shnilar usuli yoki Bayes tasnifi, to'plangan ma'lumotlar bilan ishlashning iloji yo'qligi sababli, ko'p o'lchovli ma'lumotlarga taalluqli emas.

Ma'lumotlar omborlari. Zamonaviy korxonalarning axborot tizimlari ko'pincha ma'lumotlarni kiritish va tuzatish vaqtini minimallashtiradigan tarzda tashkil etiladi, ya'ni. ma'lumotlar bazasini loyihalash nuqtai nazaridan optimal tarzda tashkil etilmagan. Ushbu yondashuv tarixiy (arxiv) ma'lumotlarga kirishni murakkablashtiradi. Axborot tizimlari ma'lumotlar bazalaridagi tuzilmalarning o'zgarishi juda mashaqqatli va ba'zan oddiygina imkonsizdir. Shu bilan birga, zamonaviy biznesni muvaffaqiyatli olib borish uchun tahlil qilish uchun qulay va real vaqtda taqdim etiladigan dolzarb ma'lumotlar talab qilinadi. Bunday

ma'lumotlarning mavjudligi ham mavjud vaziyatni baholashga, ham kelajakka bashorat qilishga imkon beradi, shuning uchun yanada muvozanatli va asosli qarorlar qabul qiladi. Bundan tashqari, haqiqiy ma'lumotlar qaror qabul qilish uchun asos bo'lishi kerak. Agar ma'lumotlar korxonaning turli xil axborot tizimlarining ma'lumotlar bazalarida saqlansa, ularni tahlil qilish paytida bir qator qiyinchiliklar yuzaga keladi, xususan, so'rovlarni ko'rib chiqish uchun vaqt sezilarli darajada oshadi; turli xil ma'lumotlar formatlarini qo'llab-quvvatlashda, shuningdek ularni kodlashda muammolar bo'lishi mumkin; tarixiy ma'lumotlarning uzoq muddatli seriyasini tahlil qilishning iloji yo'qligi va boshqalar. Ushbu muammo ma'lumotlar omborini yaratish orqali hal qilinadi. Bunday omborning vazifasi boshqarish ob’yektining yaxlit izchil ko'rinishini shakllantirish uchun heterojen manbalardan olingan operatsion ma'lumotlarni birlashtirish, yangilash va yarashtirishdir. Ma'lumotlar omborlari asosida barcha turdagi hisobotlarni tuzish, shuningdek tezkor tahliliy ishlov berish va Data Mining-ni amalga oshirish mumkin. Bill Inmon ma'lumotlar omborlarini "boshqaruvni qo'llab-quvvatlash uchun tashkil etilgan domenga xos, yaxlit, o'zgarmas, tarixiy ma'lumotlar to'plami" deb ta'riflaydi va menejerlar va tahlilchilarga ishonchli ma'lumotlarni taqdim etuvchi "haqiqatning yagona va yagona manbai" sifatida ishlab chiqilgan. va qaror qabul qilish.

Ma'lumotlar ombori tushunchasi, ma'lumotlar ulardan foydalanadigan dasturlarga emas, balki u tavsiflaydigan domenlarga qarab tasniflanadi va saqlanadi.

Integratsiya ma'lumotlar bitta biznes funktsiyasini emas, balki butun korxona ehtiyojlarini qondirishini anglatadi. Shunday qilib, ma'lumotlar ombori turli xil tahlilchilar uchun tuzilgan bir xil hisobotlarda bir xil natijalarga ega bo'lishini ta'minlaydi.

Vaqt ma'lumotnomasi shuni anglatadiki, saqlashni "tarixiy" ma'lumotlar to'plami sifatida ko'rish mumkin: ma'lumotlar istalgan vaqtgacha tiklanishi mumkin. Vaqt atributi ma'lumotlar ombori tuzilmalarida aniq mavjud.

O'zgarmaslik degani, u omborga tushgandan so'ng, ma'lumotlar omborda saqlanadi va o'zgarmaydi. Ma'lumotlar faqat xotiraga qo'shilishi mumkin.

Ushbu kontseptsiyaning yaratuvchisi Richard Xekatornning yozishicha, Ma'lumotlar omborlarining maqsadi tashkilotni "mavjud haqiqatning yagona qiyofasi" bilan ta'minlashdir. Boshqacha qilib aytganda, ma'lumotlar ombori korxona faoliyati to'g'risidagi ma'lumotlarning o'ziga xos akkumulyatoridir. Ombordagi ma'lumotlar "yulduz" yoki "qor parchasi" deb nomlangan ko'p o'lchovli tuzilmalar shaklida taqdim etilgan.

Ma'lumotlar omborlaridan foydalanishning afzalliklari. Ma'lumotlar ombori operatsion tizimlar yoki ma'lumotlar bazalarini ishlatishda afzalliklarga ega, quyidagilar keltirilgan:

- Operatsion tizimlardan farqli o'laroq, ma'lumotlar ombori bir vaqtning o'zida barcha kerakli vaqt oralig'ida - bir necha o'n yillargacha bo'lgan ma'lumotlarni bitta axborot makonida o'z ichiga oladi, bu esa bunday omborlarni tendentsiyalar, mavsumiy bog'liqliklar va boshqa muhim tahliliy ko'rsatkichlarni aniqlash uchun ideal asosga aylantiradi.
- Odatda, korporativ axborot tizimlari o'xshash ma'lumotlarni har xil usulda saqlaydi va taqdim etadi. Masalan, bir xil ko'rsatkichlar turli o'lchov birliklarida saqlanishi mumkin. Xuddi shu mahsulotlar yoki bir xil mijozlar boshqacha nomlanishi mumkin. Saqlash tizimlarida ma'lumotlar nomuvofiqligi ma'lumot to'plash va ularni yagona ma'lumotlar bazasiga singdirish bosqichida yo'q qilinadi. Shu bilan birga, barcha ko'rsatkichlar bir xil o'lchov birliklariga tushirilgan birlashtirilgan ma'lumotnomalar tashkil etilgan.
- Ko'pincha, operatsion tizimlar operatorlarning xatosi tufayli ma'lum miqdorda noto'g'ri ma'lumotlarni o'z ichiga oladi. Uni ma'lumotlar omboriga joylashtirish bosqichida ma'lumotlar oldindan qayta ishlanadi. Maxsus texnologiyadan foydalangan holda ma'lumotlar ko'rsatilgan cheklovlarga muvofiqligi tekshiriladi va kerak bo'lganda tuzatiladi (tozalanadi). Texnologiya ishonchli ma'lumotlarga asoslangan analitik hisobotlarni tuzishni va ombor ma'murini kiruvchi ma'lumotdagi xatolar to'g'risida o'z vaqtida xabardor qilishni ta'minlaydi.

- Ma'lumotlarga kirishni universallashtirish. Ma'lumotlar ombori bitta ma'lumot manbai asosida korxona faoliyati to'g'risida har qanday hisobotlarni qabul qilishning noyob imkoniyatini beradi. Bu turli xil operatsion tizimlarga kiritilgan va to'plangan ma'lumotlarni birlashtirishga va ularni osonlikcha va sodda tarzda taqqoslashga imkon beradi. Shu bilan birga, hisobotlarni yaratish jarayonida foydalanuvchi operatsion tizimlarning ma'lumotlariga kirishdagi farqlar bilan bog'liq emas.
- Analitik hisobotlarni qabul qilishni tezlashtirish. Operatsion tizimlar tomonidan taqdim etilgan vositalardan foydalangan holda hisobotlarni olish eng maqbul usul emas. Ushbu tizimlar ma'lumotlarni jamlash uchun katta vaqt sarflaydi (umumiy, o'rtacha, minimal, maksimal qiymatlarni hisoblash). Bundan tashqari, operatsion tizimning amaldagi ma'lumotlar bazasida faqat eng kerakli va so'nggi ma'lumotlar mavjud bo'lib, o'tgan davrlar uchun ma'lumotlar arxivda saqlanadi. Agar ma'lumotni arxivdan olish kerak bo'lsa, hisobotni qurish davomiyligi yana ikki yoki uch baravar ko'payadi. Shuni ham yodda tutish kerakki, operatsion tizimning serverida bir vaqtning o'zida murakkab hisobotlarni tayyorlash va ma'lumotlarni kiritish paytida kerakli ishlash ta'minlanmaydi. Bu korxona faoliyatiga katastrofik ta'sir ko'rsatishi mumkin, chunki operatorlar navbatdagi hisobot tuzilayotganda hisob-fakturalarni rasmiylashtira olmaydi, mahsulotni jo'natgan yoki qabul qilganligini qayd eta olmaydi. Ma'lumotlar ombori ushbu muammolarni hal qiladi. Birinchidan, saqlash serveri operatorlarga xalaqit bermaydi. Ikkinchidan, saqlashda batafsil ma'lumotdan tashqari, oldindan hisoblangan yig'ilgan qiymatlar ham mavjud. Uchinchidan, omborda arxivlangan ma'lumotlar har doim hisobotlarga qo'shilishi uchun mavjud. Bularning barchasi hisobotlarni tuzish vaqtini sezilarli darajada qisqartirishga va operatsion ishda muammolardan qochishga imkon beradi.
- Maxsus so'rovlarni yaratish. Ma'lumotlar omboridagi ma'lumotlarni markazlashtirish va tuzish etarli emas. Tahlilchiga ushbu ma'lumotni tasavvur qilish vositasi, o'z vaqtida qaror qabul qilish uchun zarur bo'lgan ma'lumotlarni olish oson bo'lgan vosita kerak. Har qanday tahlilchining asosiy talablaridan biri bu hisobotlarni tayyorlashning soddaligi va ularning ravshanligi. Onlayn tizimlarda

hisobot ko'pincha egiluvchan emas; yangi hisobot yaratish uchun siz bir nechta tizim ma'lumotlarini birlashtiradigan IT mutaxassislarini jalb qilishingiz kerak. Ma'lumotlar omboridan foydalanganda, masalaning echimi OLAP (On-Line Analytical Processing) texnologiyasi bilan ta'minlanadi. Ushbu texnologiya ma'lumotlarga tahlilchiga tanish bo'lgan sharoitlarda kirishni ta'minlaydi. OLAP texnologiyasi ko'p o'lchovli ma'lumotlarni taqdim etish kontseptsiyasiga asoslangan. Darhaqiqat, ma'lumotlar omborida joylashgan har bir raqamli qiymat bir necha o'nlab atributlarga ega (masalan, ma'lum bir mintaqada ma'lum bir menejer tomonidan ma'lum bir sana bo'yicha sotishlar soni va boshqalar). Shunday qilib, ish raqamli qiymatlar bir necha o'lchovlar kesishmasida joylashgan ko'p o'lchovli ma'lumotlar tuzilmalari (ko'p o'lchovli kublar) bilan amalga oshiriladi deb taxmin qilishimiz mumkin. Bu OLAP tizimlarida qo'llaniladigan yondashuv. Ular OLAP manipulyatsiyasi deb ataladigan ko'p o'lchovli tuzilmalar bo'ylab harakatlanishning moslashuvchan vositalarini taqdim etadilar. Ularning yordami bilan tahlilchi turli xil bo'laklarni qabul qilishi, ma'lumotlarni "burish" mumkin.

Ma'lumotlarni saqlash texnologiyasidan foydalanishning sanab o'tilgan afzalliklaridan ko'rinib turibdiki, ularning aksariyati Data Mining jarayonini sezilarli darajada soddalashtirishi, tezligini oshirishi va sifat jihatidan yaxshilashi mumkin. Shunday qilib, ushbu texnologiyalarni kompleks tatbiq etilishi ishlab chiquvchilar va foydalanuvchilarga qarorlarni qo'llab-quvvatlash tizimlarini yaratishda turli xil axborot tizimlarining turlicha ma'lumotlar bazalaridan foydalanishiga nisbatan inkor etib bo'lmaydigan afzalliklarni beradi.

Nazorat savollari:

1. Ma`lumot saqlagich deganda nimani tushunasiz?
2. OLAP tizimlariga misol keltiring?
3. OLAP, MOLAP,HOLAP tizimlarining farqi nimada?
4. Integratsiya nima? Misol keltiring.

GLOSSARIY

Автоматизация **Automation** **Avtomatlasshtirish**	это применение в производстве технических средств, методов и систем управления, освобождающих человека от непосредственного участия в производстве.	this application in the production of technical means, methods and control systems that relieve a person from direct participation in production.	Ushbu qo'llanmani ishlab chiqarishda bevosita ishtirok etishdan ozod qiluvchi texnik vositalar, usullar va nazorat qilish tizimlarini ishlab chiqarish.
САПР **CAD** **ALS**	(англ. CAD, Computer-Aided Design) - программный пакет, предназначенный для проектирования (разработки) объектов производства (или строительства), а также оформления конструкторской и/или технологической документации	(English CAD, Computer-Aided Design) - a software package designed for the design (development) of production facilities (or construction), as well as design and / or technological documentation.	(Ingliz CAD, Computer Assisted Design) - ishlab chiqarish ob'ektlarini (yoki qurilishi), shuningdek loyihalashtirish va / yoki texnologik hujjatlarni loyihalash uchun ishlab chiqilgan dasturiy ta'minot to'plami.
MATLAB MATLAB MATLAB	(англ. Matrix Laboratory) – среда высокой производительности, предназначенная для технических вычислений, их интегрирования, визуализации и программирование в удобной для использования среде, где задачи и решения выражаются в привычной математической нотации.	(English Matrix Laboratory) is a high performance environment designed for technical computing, integrating, visualizing and programming in an easy-to-use environment where tasks and solutions are expressed in familiar mathematical notation.	(Ingliz Matritsa laboratoriyasi) texnik hisoblash, integratsiya qilish, tasavvur qilish va programmalarga moslashtirilgan qulay muhitda dasturlash uchun mo'ljallangan yuqori ishlash muhiti bo'lib, u erda vazifalar va echimlar tanish matematik notada ifodalangan.
SCADA **SCADA** **SCADA**	сокр. от Supervisory Control And Data Acquisition диспетчерское управление и сбор данных, (название класса систем для комплексной автоматизации промышленного производства)	abbr. from Supervisory Control And Data Acquisition dispatching management and data collection, (name of a class of systems for integrated automation of industrial production)	Abbr. Kuzatuv nazorati va ma'lumotlarni to'plash dispetcherlik boshqaruvi va ma'lumotlar yig'ishdan (sanoat ishlab chiqarishni kompleks avtomatlashtirish uchun tizimlar sinfining nomi)
TRACE MODE **TRACE MODE** **TRACE MODE**	первая интегрированная информационная система для управления промышленным производством, объединяющая в едином целом продукты класса SOFTLOGIC-SCADA / HMI-MESEAM-HRM)	the first integrated information system for industrial production management, which unites SOFTLOGIC-SCADA / HMIMES-EAM-HRM products in a single whole)	SOFTLOGIC-SCADA / HMI-MES-EAM-HRM mahsulotlarini bir butunda birlashtirgan sanoatni ishlab chiqarishni boshqarish bo'yicha birinchi integratsiya axborot tizimi
SIMATIC WinCC **SIMATIC WinCC**	(Windows Control Center) - это компьютерная система	(Windows Control Center) is a computer-	(Windows Boshqarish Markazlari) Windows

SIMATIC WinCC	человекомашинного интерфейса, работающая под управлением операционных систем Windows и предоставляющая широкие функциональные возможности для построения систем управления различного назначения и уровней автоматизации	based human-machine interface system running under Windows operating systems and providing extensive functionality for building management systems for various purposes and automation levels	operatsion tizimlari ostida ishlaydigan va turli maqsadlar uchun boshqaruv tizimlari va avtomatlashtirish darajalari uchun keng funksionallikni ta'minlaydigan kompyuterga asoslangan inson-mashina interfeysi tizimidir
Автоматизированная Обучающая Система (АОС) Automated Training System Avtomatlashtirilgan o‘quv tizimlari	программное средство профессиональной подготовки персонала, состоящее из одного или нескольких автоматизированных учебных курсов (АУК) и набора специализированных локальных тренажеров, позволяющих осуществлять формирование профессиональных навыков и умений принятия и выполнения решений по управлению (обслуживанию) объектов, рассматриваемых в содержательной части АУК	software for professional training of personnel consisting of one or more automated training courses (AUC) and a set of specialized local simulators that allow the formation of professional skills and skills in the adoption and implementation of decisions on management (maintenance) of objects considered in the content part of the AUC.	avtomatlashtirilgan o'quv kurslarida (AUC) va professional ko'nikmalarini qabul qilinishi va boshqaruv yechimlari (xizmat) ob'ekt amalga oshirish shakllantirish uchun imkon ixtisoslashgan mahalliy murabbiylar bir qator bir yoki undan ortiq iborat kasbiy ta'lim, dasturiy vositalari, AUC mazmunini ko'rib chiqildi.
Автоматизированны й Учебный Курс (АУК) Automated Training Course Avtomatlashtirilgan o‘quv kursi	программное средство профессиональной подготовки персонала, отвечающее требованиям методик подготовки, реализующее предъявление обучаемому графического и текстового материала нормативнотехнической документации конкретного учебного курса и обеспечивающее контроль качества подготовки обучаемых	software for professional training of personnel that meets the requirements of training methods that implements the presentation of the graphic and text material to the trainee of the normative and technical documentation of the specific training course and ensures the quality control of the trainees	Kadrlarni malakali o'qitish uchun maxsus o'quv kursining normativ-texnik hujjatlarini tinglovchilarga grafik va matn materiallarini taqdim etishni o'rgatadigan o'qitish metodlari talablariga javob beradigan dasturiy ta'minot va malaka oshirish kurslarining sifatini nazorat qilish
Автоном-ный тренажер Independent trainer Avtonom trenajor	тренажер оператора системы «человек-машина», функционирующий без системы «человек-машина»	The trainer of the operator of the "man-machine" system, functioning without the "manmachine" system	"man-mashina" tizimi bo'lmagan holda ishlaydigan "man-mashin" tizimi operatorining murabbiysi
Адаптив-ный тренажер Adaptive	тренажер оператора системы «человек-машина»,	the trainer of the operator of the "man-machine" system, which provides	"Man-mashina" tizimining operatorini o'qitish jarayonini

simulator Adaptiv trenajor	обеспечивающий автоматическую оптимизацию управления процессом подготовки оператора системы «человек-машина» с учетом результатов выполнения им учебных задач	automatic optimization of the management of the process of training the operator of the "man-machine" system, taking into account the results of the performance of the training tasks	boshqarishning avtomatlashtirilgan optimallashini ta'minlaydigan "manmashina" tizimining operatori o'qituvchisi, mashg'ulot vazifalarini bajarish natijalarini hisobga olgan holda
Встроенный тренажер Built-in simulator Ajralmas trenajor	тренажер оператора системы «человек-машина», функционирующий совместно с системой «человек-машина»	The simulator of the operator of the "man-machine" system, functioning in conjunction with the "man-machine" system	"inson mashinasi" tizimi bilan ishlaydigan "insonmashina" tizimining operatori simulyatori
Групповой тренажер Group simulator Guruhli trenajor	тренажер оператора системы «человек-машина», предназначенный для одновременной подготовки операторов взаимосвязанных систем «человек-машина»	the trainer of the operator of the system "man-machine", intended for simultaneous preparation of operators of the interconnected systems "personmachine"	"Man-mashina" tizimining operatorlarini bir vaqtning o'zida ishlab chiqarishga mo'ljallangan "odammashin" tizimining operatori,
Информа-ционная модель системы «человек-машина» Information model of "man-machine" system «Inson-mashina» tizimining informatik modeli	условное отображение информации о состоянии объекта воздействия, системы «человекмашина» и способов управления ими	Conditional display of information on the state of the impact object, the "manmachine" system and the ways to manage them	ta'sir ob'ekti holati, "odammashina" tizimi va ularni boshqarish usullari to'g'risida shartli ma'lumotlarni ko'rsatish
Надежность оператора системы «человек-машина» The reliability of the system operator "man-machine" «Inson-mashina» operator tizimining ishonchliligi	свойство человека-оператора системы «человек-машина» сохранять работоспособное состояние в течение требуемого интервала времени	The property of the personoperator of the "man-machine" system to maintain an operational state for the required time interval	"Man-mashina" tizimining shaxs-operatorining mulki talab qilinadigan vaqt oralig'ida operatsion holatni saqlab qolish uchun
Citect SCADA Citect SCADA Citect SCADA	– программный продукт, представляющий собой полнофункциональную систему визуализации и мониторинга, управления и сбора данных.	- software product, which is a full-featured system of visualization and monitoring, management and data collection.	Vizualizatsiya qilish va monitoring qilish, boshqarish va ma'lumotlar yig'ishning to'liq xususiyatli dasturiy mahsuloti.
SCADA система InTouch InTouch SCADA system SCADA InTouch tizimi	– это достаточно мощная среда разработки визуализации и управления для промышленной автоматизации технологических процессов и диспетчерского контроля,	Is a powerful visualization and control environment for industrial automation of technological processes and dispatch control, it is used to create DCS (distributed control	Texnologik jarayonlar va dispetcherlik boshqaruvini sanoat avtomatizatsiyasi uchun kuchli vizualizatsiya va boshqarish muhiti, DCS (tarqalgan boshqaruv tizimlari) va

	применяется для создания DCS (распределенных систем управления) и других АСУ ТП.	systems) and other process control systems.	boshqa jarayonlarni boshqarish tizimlarini yaratish uchun ishlatiladi.
Системное ПО System Software Tizimili dasturiy ta’minot	– совокупность программ, обеспечивающих общее управление функционированием вычислительной системы и выполнение функций по ее обслуживанию.	- a set of programs that provide overall management of the functioning of the computer system and the performance of the functions for its maintenance.	- kompyuter tizimining ishlashi ustidan umumiy nazoratni ta'minlaydigan va unga xizmat ko'rsatish funktsiyalarini bajaradigan dasturlarning to'plami
Инструментальное ПО Tool Software Uskunali dasturiy ta’minot	- совокупность средств, обеспечивающих процесс функционирования прикладных программ, т.е. обеспечивающих перевод кода программ, написанных пользователем с помощью высокоуровневых языком программирования на языки более низкого уровня – в машинный код.	- a set of tools that provide the process of functioning of application programs, i.e. providing the translation of the code of programs written by the user with the help of high-level programming language to languages of a lower level - into machine code.	- dastur dasturlarining ishlashini ta'minlaydigan vositalar majmuasi, ya'ni. foydalanuvchi tomonidan yuqori darajadagi dasturlash tilining past darajadagi tillarga tarjima qilingan kodlari tarjimasini mashina kodiga tarjima qilish.

ASOSIY 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-Петерберг, 2004. - 336 с.
2. Елманова Н., Федоров А. Введение в OLAP-технологии Microsoft. СПб.: БХВ-Петерберг, 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 .ComputerWeek-Москва. 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
11. Медведев В.С., Потемкин В.Г. Нейронные сети. Matlab 6. Диалог-МИФИ. 2002, 496 стр
12. Э.А. Тахтенгерц. Компьютерная поддержка принятия решений. М.: СИНТЕГ. 1998, с. 376, с

INTERNET SAYTLARI

http://ziyonet.uz - Milliy ijtimoiy-ta`lim tarmog`i

http://www.rsl.ru - Rossiya davlat kutubxonasi

https://basegroup.ru/ - BaseGroup kompaniyasi sayti

https://www.sap.com/ - KXEN kompaniyasi sayti