

14-§. Qaror daraxtini qurish usullari.

Reja:

1. *Qaror daraxtini qurish usullari va afzalliklari*
2. *Qarorlar daraxtining foydalanish sohalari*

Qaror daraxtlari - ma'lumotlarni avtomatik tahlil qilish usullaridan biridir. Biz ishning umumiy tamoyillarini va qo'llash sohalarini tahlil qilamiz.

Qaror daraxtlari ma'lumotlarni avtomatik tahlil qilish usullaridan biridir. Qaror daraxtlari tasniflash va regressiya muammolarini hal qilish uchun eng samarali ma'lumotlarni yig'ish va bashoratli tahlil vositalaridan biridir. Qaror daraxtlari tasniflash va regressiya muammolarini hal qilish uchun eng kuchli ma'lumotlarni qidirish va bashoratli tahlil vositalaridan biridir.

Ular "Agar ..., keyin ..." shaklidagi qaror qoidalaridan tashkil topgan ierarxik daraxt tuzilmalaridir. Qoidalar o'quv majmuasi bo'yicha mashg'ulotlar paytida avtomatik ravishda yaratiladi va ular deyarli tabiiy tilda tuzilganligi sababli (masalan, "Agar savdo hajmi 1000 donadan ortiq bo'lsa, u holda mahsulot istiqbolli"), analitik modellar sifatida qaror daraxtlari, aytaylik, neyron tarmoqlarga

qaraganda ko'proq og'zaki va izohlanadi.

Qaror daraxtlarining tuzilishi:

Sarlavha	Tavsif
Ob'ekt	misol, naqsh, kuzatish
Maqsadli o'zgaruvchi	bog'liq o'zgaruvchi, sinf belgisi
Tugun	ichki daraxt tugun, tekshirish tugun
Ildiz tugun	qarorlar daraxtining boshlang'ich tuguni

Barg	daraxt oxiri tugun, qaror tugun, terminal tugun
Qaror qabul qilish qoidasi	tugundagi holat, tekshirish

Qaror daraxtlaridagi qoidalar mavzu sohasini tavsiflovchi ko'plab individual kuzatishlarni (o'qitish misollarini) umumlashtirish yo'li bilan olinganligi sababli, mantiqiy xulosa chiqarishning tegishli usuliga o'xshab, ular induktiv qoidalar, o'quv jarayonining o'zi esa qarorlar daraxtlarining induksiyasi deb ataladi.

Trening to'plamida misollar uchun maqsadli qiymat ko'rsatilishi kerak, chunki qaror daraxtlari - bu nazorat ostida o'rganish yordamida qurilgan modellar. Bundan tashqari, agar maqsadli o'zgaruvchi diskret (sinf yorlig'i) bo'lsa, u holda model tasniflash daraxti deb ataladi va agar u uzluksiz bo'lsa, regressiya daraxti deb ataladi.

Qaror daraxtlarining asosiy qo'llanilishi statistika, ma'lumotlarni tahlil qilish va mashinani o'rganishda qo'llaniladigan boshqaruv qarorlarini qabul qilish jarayonlarini qo'llab-quvvatlashdir. Ushbu qurilma yordamida hal qilinadigan vazifalar:

Tasniflash - ob'ektlarni oldindan ma'lum bo'lgan sinflardan biriga belgilash. Maqsadli o'zgaruvchi diskret qiymatlarga ega bo'lishi kerak.

Regressiya (raqamli bashorat) - berilgan kirish vektori uchun mustaqil o'zgaruvchining raqamli qiymatini bashorat qilish.

Ob'ektlar tavsifi - qarorlar daraxtidagi qoidalar to'plami ob'ektlarni ixcham tasvirlash imkonini beradi. Shuning uchun, ob'ektlarni tavsiflovchi murakkab tuzilmalar o'rniga siz qaror daraxtlarini saqlashingiz mumkin.

Qaror daraxtlarining paydo bo'lishi va rivojlanishiga fundamental g'oyalar 1950-yillarda kompyuter tizimlari yordamida inson xatti-harakatlarini modellashtirish bo'yicha tadqiqotlar sohasida boshlangan. Ular orasida

K.Xovlendning “Tafakkurni kompyuter modellashtirish” va E.Hant va boshqalarning “Induksion tajribalar” asarlarini alohida ta’kidlash lozim.

Ma’lumotlarni tahlil qilish uchun o’z-o’zini o’rganish modellari sifatida qarorlar daraxtlarining keyingi rivojlanishi ID3 algoritmini va uning takomillashtirilgan C4.5 va C5.0 modifikatsiyalarini ishlab chiqqan Jon R. Quinlan, shuningdek Leo Breiman nomlari bilan bog’liq, CART algoritmini va tasodifiy tanlash taklif qilgan.

Hozirgi vaqtda ko’plab qarorlar daraxtini o’rganish algoritmlari ishlab chiqilgan:

ID3, CART, C4.5, C5.0, NewId, ITrule, CHAID, CN2 va boshqalar. Ammo eng keng tarqalgan va mashhurlari quyidagilardir:

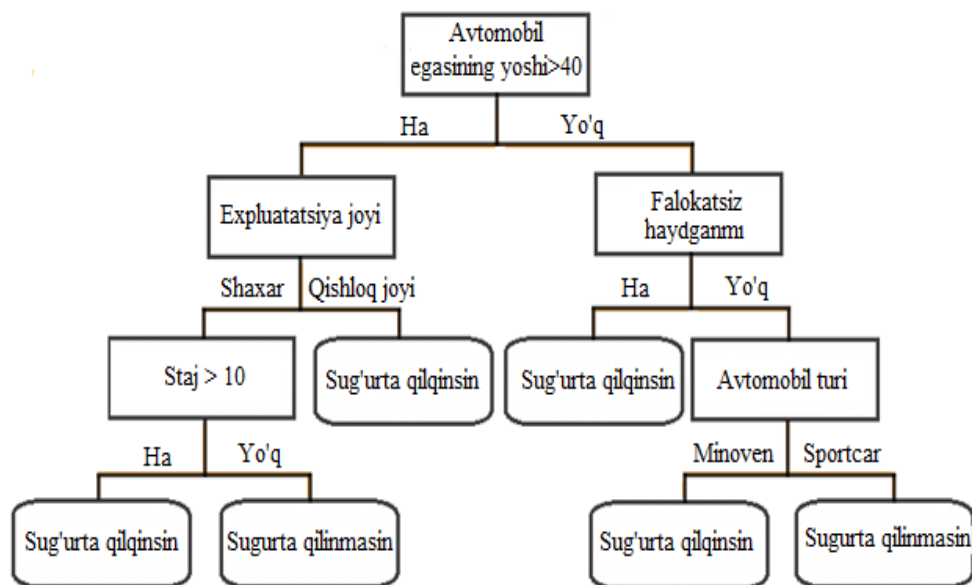
ID3 (Iterative Dichotomizer 3) - algoritm faqat diskret maqsadli o’zgaruvchi bilan ishlashga imkon beradi, shuning uchun ushbu algoritm yordamida qurilgan qaror daraxtlari tasniflanadi. Daraxt tugunidagi barglar soni cheklanmagan. Yo’qolgan ma’lumotlarni qayta ishlash mumkin emas.

C4.5 - ID3 algoritmining takomillashtirilgan versiyasi bo’lib, u etishmayotgan atribut qiymatlari bilan ishlash qobiliyatini qo’shadi.

CART (Tasniflash va regressiya daraxti) qarorlar daraxtlarini o’rgatish algoritmi bo’lib, u ham diskret, ham doimiy maqsadli o’zgaruvchilardan foydalanishga, ya’ni tasniflash va regressiya masalalarini hal qilishga imkon beradi. Algoritm har bir tugunda faqat ikkita bargga ega bo’lgan daraxtlarni quradi.

Qaror daraxtlaridagi qoidalar mavzu sohasini tavsiflovchi individual kuzatishlar (o’quv misollari) to’plamini umumlashtirish orqali olinganligi sababli, tegishli xulosa chiqarish usuliga o’xshab, ular induktiv qoidalar deb ataladi va o’quv jarayonining o’zi qaror daraxt induksiyasi deb ataladi.

Qarorlar daraxtiga misol

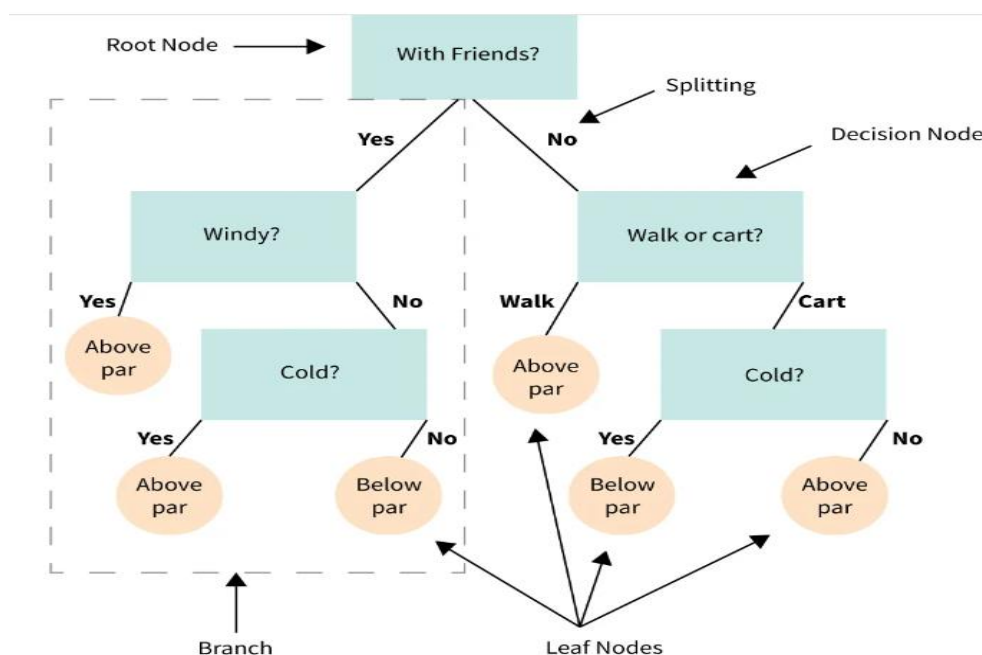


Trening to'plamida misollar maqsadli qiymatga ega bo'lishi kerak, masalan qaror daraxtlari - bu nazorat ostida o'rganish asosida qurilgan modellar. Bunday holda, agar maqsadli o'zgaruvchi diskret bo'lsa, u holda model tasniflash daraxti, doimiy bo'lsa, regressiya daraxti deb ataladi.

Qaror daraxtlarining paydo bo'lishi va rivojlanishiga turtki bo'lgan fundamental g'oyalar 1950-yillarda kompyuter tizimlari yordamida inson xatti-harakatlarini modellashtirish bo'yicha tadqiqotlar sohasida yaratilgan. Ular orasida K.Xovlend ning "Tafakkurni kompyuter modellashtirishi" va E.Hant va boshqalarning "Induksiya bo'yicha eksperimentlar" asarlari bor.

Ma'lumotlarni tahlil qilish uchun o'z-o'zini o'rganish modellari sifatida qarorlar daraxtlarining keyingi rivojlanishi ID3 algoritmini va uning takomillashtirilgan C4.5 va C5.0 modifikatsiyalarini ishlab chiqqan Jon R. Quinlan, shuningdek, Leo Breiman CART algoritmi va tasodifiy o'rmon usuli algoritm nomlari bilan bog'liq. Qarorlar daraxtlari nazariyasida qo'llaniladigan asosiy tushunchalar bilan tanishamiz:

Qaror daraxtlari va uni qurish algoritmi:



Aslida, qarorlar daraxtining o'zi ikki turdagi elementlardan tashkil topgan ierarxik tuzilishda qaror qoidalarini ifodalash usuli hisoblanadi –tugunlar (node) va barglar (leaf) tugunlar qaror qoidalarini o'z ichiga oladi va misollarning ushbu qoidaga muvofiqligi o'quv majmuasining ba'zi bir atributi uchun tekshiriladi.

Eng oddiy holatda, tekshirish natijasida tugunga tushadigan misollar to'plami ikkita kichik to'plamga bo'linadi, ulardan birida qoidani qanoatlantiradigan, ikkinchisida esa mos kelmaydigan misollar mavjud.

Keyin, qoida har bir kichik to'plamga yana qo'llaniladi va algoritmnining qandaydir to'xtash shartiga erishilgunga qadar protsedura rekursiv ravishda takrorlanadi. Natijada, oxirgi tugun tekshirilmaydi va bo'linmaydi va u barg deb e'lon qilinadi. Varraq unga tushadigan har bir misol uchun yechimni aniqlaydi. Tasniflash daraxti uchun bu tugun bilan bog'langan sinf, regressiya daraxti uchun esa bargga mos keladigan maqsadli o'zgaruvchining modal oralig'idir.

Shunday qilib, tugundan farqli o'laroq, barg qoidani o'z ichiga olmaydi, lekin bu barg bilan tugaydigan filialning barcha qoidalarini qondiradigan ob'ektlarning kichik to'plamidir. Shubhasiz, bargga kirish uchun misol bu bargga boradigan yo'lda yotadigan barcha qoidalarni qondirishi kerak. Daraxtdagi har

bir bargga boradigan yo'l noyob bo'lgani uchun, har bir misol faqat bitta bargga kirishi mumkin, bu esa yechimning o'ziga xosligini ta'minlaydi.

Qarorlar daraxti qurilishining asosiy bosqichlari

Qarorlar daraxtini yaratishda bir nechta asosiy muammolarni hal qilish kerak, ularning har biri o'quv jarayonining tegishli bosqichi bilan bog'liq:

- Ushbu tugunda bo'linish amalga oshiriladigan atributni tanlash (split atributi).
- O'rganishni to'xtatish mezonini tanlash.
- Filiallarni kesish usulini tanlash (soddalashtirish).
- Qurilgan daraxtning aniqligini baholash.

Qaror qabul qilishning afzalliklari:

PROS	MUHIMLAR
<ul style="list-style-type: none">✓ Bu ma'lumotlarni sharhlash uchun ajoyib vositadir.✓ Bu raqamli va raqamli bo'lmagan ma'lumotlarni qayta ishlash uchun javob beradi.✓ Uni yaratish yoki ishlatishdan oldin u faqat minimal tayyorgarlikni talab qiladi.✓ Bu eng yaxshi, eng yomon va eng ehtimoliy stsenariylarni tanlashni osonlashtiradi.✓ Siz qaror daraxtlarini boshqa qaror qabul qilish usullari bilan osongina birlashtira olasiz.	<ul style="list-style-type: none">✗ Agar qaror daraxtining dizayni juda murakkab bo'lsa, ortiqcha moslama paydo bo'lishi mumkin. Va bu ko'plab foydalanuvchilar uchun muammoga aylanadi.✗ Qaror daraxtlari doimiy o'zgaruvchilar uchun mos emas (bir nechta qiymatga ega o'zgaruvchilar).✗ Bashoratli tahlilga kelganda, hisob-kitoblar noqulay bo'lishi mumkin.✗ Qaror daraxtlari boshqa bashorat qilish usullariga nisbatan pastroq bashorat aniqligini ishlab chiqaradi.

Qaror daraxtini qurishda bir nechta asosiy muammolarni hal qilish kerak, ularning har biri ma'lumotlarni qarorni qabul qilishda jarayonining tegishli bosqichi bilan bog'liq:

- Tugunda bo'linish amalga oshiriladigan atributni tanlash (bo'lim atributi).
- Qaror daraxti mezonini tanlash.
- Qaror daraxtini usulini tanlash (soddalashtirish).
- Qurilgan daraxtning to'g'riligini baholash.

Qaror daraxtini qurishning afzalliklar:

- Raqamli yoki toifali ma'lumotlar va o'zgaruvchilar uchun ishlaydi.
- Bir nechta chiqishlar bilan muammolarni modellashtiradi.

- Daraxtning ishonchliligini tekshiradi.
- Boshqa ma'lumotlarni modellashtirish usullariga qaraganda kamroq ma'lumotlarni talab qiladi.
- Analitik ma'lumotga ega bo'lmaganlarga tushuntirish oson.

Qaror daraxtini qurishning kamchiliklari:

- Ma'lumotlarning ortiqcha ta'svirlanganligi.
- Katta ma'lumotlar to'plamlari uchun ideal emas.
- Atributlarni nomutanosib ravishda baholashi mavjud emas.
- Tugunlardagi qarorlar ikkilik natijalar bilan cheklangan bo'lib, daraxt hal qila oladigan darajada murakkablashishi.
- Noaniqlik va ko'plab bog'liq natijalar bilan shug'ullanganda daraxtlar juda murakkab bo'lishi mumkin.

Qarorlar daraxtining foydalanish sohalari

- Qaror daraxtlarini yaratish va o'rganish uchun modullar ko'pchilik analitik platformalarga kiritilgan. Ular qarorlarni qo'llab-quvvatlash tizimlari va ma'lumotlarni ishlab chiqishda qulay vositadir.
- Qaror daraxtlari amalda quyidagi sohalarda muvaffaqiyatli qo'llaniladi:
- Bank ishi - Kredit berishda bank mijozlarining kreditga layoqatliligini baholash.
- Sanoat - Mahsulot sifatini nazorat qilish (nuqsonlarni aniqlash), buzilmaydigan sinov (masalan, payvandlash sifatini tekshirish) va boshqalar.
- Tibbiyot - Kasalliklarning diagnostikasi.
- Molekulyar biologiya - Aminokislotalarning tuzilishini tahlil qilish.
- Savdo - Mijozlar va tovarlarni tasniflash.

Nazorat savollari.

- .

Qaror daraxtini qurish usullari va afzalliklarini izohlang.

- .

Qarorlar daraxtining foydalanish sohalarini izohlang.



Mavzuni mustahkamlash uchun savollar.

1.Qaror daraxtlari tabiiy tilda tuzilishi mumkin bo'lgan qanday shaklidagi qaror qoidalariga asoslanadi?

- a) "agar ... keyin ..."
- b) "rost... bo'lsa"
- c) " mantiqiy amallarga"
- d) "rost... inkor"

2.Quyidagi ta'riflardan qaysi biri BigData-ga taalluqli emas?

- a) Yuqoridagilarning barchasi taalluqli
- b) Katta hajmdagi ma'lumotlarni saqlash va qayta ishlash muammosi
- c) Katta hajmdagi ma'lumotlarni qayta ishlash usullarining kompleksto'plami
- d) Ma'lumotlar manbalarining yuqori o'zgaruvchanligi va munosabatlarning murakkabligi bilan bog'liq ma'lumotlar;

3.Muayyan hodisaning yuzaga kelish ehtimolini bashorat qilish uchun modelni yaratish va tanlash jarayoni nima deb nomlanadi?

- a) Bashoratli modellashtirish
- b) OLAP
- c) Ma'lumotlarni qazib olish
- d) Ma'lumotlar ilmi