

Qarorlar daraxtini qurish algoritmlari**QARORLAR DARAXTINI QURISH ALGORITMLARI****Onarqulov Maqsadjon Karimberdiyevich***Farg'ona davlat universiteti amaliy matematika va informatika**kafedrasi dotsenti maxmaqsad@gmail.com***Yusupov Mirsaid Abdulaziz o'g'li***Farg'ona davlat universiteti amaliy matematika va informatika**kafedrasi o'qituvchisi mirsaidbeky@gmail.com***Xudoyberdiyev Nozimbek Zarifjon o'g'li***Farg'ona davlat universiteti amaliy matematika yo'nalishi 3-bosqich talabasi**nozimbekxudoyberdiyev55@gmail.com*

Annotasiya: Ushbu maqolada qarorlar daraxti algoritmlari haqida tushuncha beriladi. Qarorlar daraxtlari ma'lumotlarni klassifikatsiya va regressiya uchun foydalaniladigan mashina o'qitishning asosiy usullaridan biri sifatida ko'rib chiqiladi. Shuningdek, qarorlar daraxtini qurish jarayonining bosqichlari va amaliy qo'llanilish sohalari ko'rsatiladi. Mazkur tadqiqot algoritmlarning nazariy jihatlarini yoritib, ularning real hayotdagi qo'llanilish imkoniyatlarini ochib beradi.

Kalit so'zlar: Qarorlar daraxti, klassifikatsiya, regressiya, vizual, interpretatsiya

Аннотация: В этой статье представлено введение в алгоритмы дерева решений. Деревья решений рассматриваются как один из основных методов машинного обучения, используемых для классификации и регрессии данных. Также показаны этапы процесса построения дерева решений и области практического применения. В этом исследовании освещаются теоретические аспекты алгоритмов и раскрываются их возможности для практического применения.

Ключевые слова: Дерево решений, классификация, регрессия, визуализация, интерпретация.

Annotation: This article provides an introduction to decision tree algorithms. Decision trees are considered as one of the main methods of machine learning used for data classification and regression. Also, the stages of the decision tree construction process and areas of practical application are shown. This study highlights the theoretical aspects of algorithms and reveals their possibilities for real-life application.

Key words: Decision tree, classification, regression, visualization, interpretation.

Qarolar daraxti nima?

Qarorlar daraxti -ma'lumotlarni avtomatik tahlil qilish usullaridan biridir. Biz ishning umumiy tamoyillarini va qo'llash sohalarini tahlil qilamiz. Qaror daraxtlari tasniflashva regressiya muammolarini hal qilish uchun eng kuchli ma'lumotlarni qidirishva 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 tabiiytilda 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 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 induksiya sideb ataladi. 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 (sinf yorlig'i), 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.Xovlendning "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 algoritmining uning takomillashtirilgan C4.5 va C5.0 modifikatsiyalarini ishlab chiqqan Jon R. Quinlan, shuningdek, Leo Breiman CART algoritmi va tasodifiy o'rmon usuli algoritmlari bilan bog'liq.

Qarorlar daraxti tuzilishi: 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. Varaq 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 qoidalarini 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.

Qaror daraxti vazifalari: Qaror daraxtlarining asosiy doirasi 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 tuzilmalarni siz qaror daraxtlarini saqlashingiz mumkin.

Qarorlar daraxtining asosiy afzalliklari:

Ma'lumotlarning vizual tahlili oson.

To'g'ridan-to'g'ri interpretatsiya qilish mumkin.

Raqamli va kategorik xususiyatlar bilan ishlash imkoniyati mavjud.

Qarorlar Daraxtini Qurish Jarayoni

Qarorlar daraxtini qurish quyidagi asosiy bosqichlarni o'z ichiga oladi:

1. Dastlabki tugun yaratish: Ma'lumotlar to'plami yagona tugunga joylashtiriladi.

2. Bo'linish mezonini tanlash: Bo'linish uchun axborot ortishi (Information Gain),

Gini indeksi yoki boshqa mezonlardan foydalaniladi.

3. Qayta bo'lish: Har bir yangi tugun uchun bo'linish jarayoni takrorlanadi.

4. To'xtash shartlarini tekshirish: Tugunlarni bo'linishi tugatiladi, agar:

Har bir tugun faqat bitta sinfni aks ettirsa;

Bo'linadigan xususiyatlar qolmasa.

5. Qirqish: Ortiqcha murakkablikni oldini olish uchun daraxt qisqartiriladi.

Qarorlar daraxtining asosiy kamchiliklari:

1. Ortiqcha moslashuv (Overfitting):

Juda chuqur daraxtlar ma'lumotlarning shaxsiy holatlariga moslashib, yangi ma'lumotlarda noto'g'ri natija berishi mumkin.

2. Kam moslashuv (Underfitting):

Juda sayoz daraxtlar murakkab munosabatlarni o'rgana olmaydi, bu esa past aniqlikka olib keladi.

3. Ma'lumotlarga sezgirlik:

Ma'lumotlardagi kichik o'zgarishlar butun daraxt tuzilishini o'zgartirishi mumkin, bu esa modelning barqarorligini pasaytiradi.

4. Muvozanatsiz sinflar bilan ishlash qiyinligi:

Sinflar orasidagi nomutanosiblik yuqori bo'lsa, kam uchraydigan sinf e'tibordan chetda qolishi mumkin.

5. Optimallashtirish samaradorligi:

Katta hajmdagi ma'lumotlar uchun daraxtni qurish ko'p vaqt va resurs talab qilishi mumkin.

6. Chiziqli bo'linish cheklovi:

Daraxtlar mantiqiy chegaralarni faqat gorizontaal yoki vertikal bo'linish orqali hosil qiladi, bu murakkab chegaralarni yaxshi ifodalay olmaydi.

7. Interpretatsiya chuqurligi:

Katta daraxtlar murakkab va qiyin tushuniladigan bo'lib qolishi mumkin.

Bu kamchiliklarni kamaytirish uchun Random Forest, Bagging, yoki Gradient Boosting kabi ansambl usullari keng qo'llaniladi.

Qarorlar daraxtini qurish algoritmlari: Qaror daraxtlarini qurish uchun ko'plab zamonaviy algoritmlarga asoslanadi. Shubhasiz, ushbu texnikadan foydalanganda, qaror daraxti yuqoridan pastgacha (ildiz tugunidan barglargacha) quriladi. Hozirgi vaqtda qarorlar daraxtini o'rganish algoritmlarining muhim soni 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 tugunlar soni cheklanmagan. Yo'qolgan ma'lumotlar bilan ishlash mumkin emas. C4.5-ID3 algoritmining takomillashtirilgan versiyasi bo'lib, u yetishmayotgan atribut qiymatlari bilan ishlash qobiliyatini qo'shadi (Springer Science ma'lumotlariga ko'ra, 2008 yilda algoritm eng mashhur 10 ta Data Mining algoritmlarida 1-o'rinni egallagan). CART (Classification and Regression Tree) (Tasniflash va regressiya daraxti) -qarorlar daraxtini o'rganish algoritmi bo'lib, u ham diskret, ham doimiy maqsadli o'zgaruvchidan foydalanishga, ya'ni tasniflash va regressiya masalalarini hal qilishga imkon beradi. Algoritm har bir tugunda faqat ikkita tugunga ega bo'lgan daraxtlarni quradi. 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:

1. Ushbu tugunda bo'linish amalga oshiriladigan atributni tanlash (split atributi).

2. O'rganishni to'xtatish mezonini tanlash.

3. Filiallarni kesish usulini tanlash (soddalashtirish).

4. Qurilgan daraxtning aniqligini baholash.

Quyida qarorlar daraxti yaratish uchun eng ko'p qo'llaniladigan algoritmlar va ularning ishlash tamoyillari keltiriladi.

1. ID3 (Iterative Dichotomiser 3)

ID3 algoritmi axborot ortishi mezonidan foydalanib, har bir xususiyat uchun eng katta informatsion ajratishni (information gain) hisoblaydi. Soddaligi tufayli kichik

ma'lumotlar to'plamlari bilan samarali ishlaydi. Ammo raqamli qiymatlar bilan ishlashda cheklovlarga ega.

2. C4.5

ID3 algoritmini rivojlantirgan holda, C4.5 raqamli xususiyatlar, yo'q qiymatlar bilan ishlash va daraxtni qirqish imkoniyatini beradi. Axborot ortishining nisbiy mezonidan (gain ratio) foydalanadi.

3. CART (Classification and Regression Tree)

CART algoritmi tasniflash va regressiya uchun moslashgan. Bo'linish mezon sifatida Gini indeksi yoki dispersiya (variance reduction) ishlatiladi. Natijada ikkilik (binary) bo'linish hosil qiladi.

4. CHAID (Chi-squared Automatic Interaction Detector)

CHAID algoritmi bo'linish jarayonida chi-kvadrat statistikasi asosida ishlaydi. Bu ko'p sinfli muammolar uchun samarali.

5. Random Forest

Random Forest bir nechta qarorlar daraxtidan tashkil topgan ansambl algoritmidir. Tasodifiy tanlangan namuna va xususiyatlar asosida turli daraxtlar qurilib, ularning natijalari birlashtiriladi. Bu usul overfittingni kamaytiradi va barqaror natijalar beradi.

6. XGBoost

XGBoost gradient boostingi asosida ishlaydi. Katta hajmdagi ma'lumotlar bilan ishlashda yuqori aniqlik va samaradorlikni ta'minlaydi.

Qarorlar daraxtlari mashina o'qitishdagi asosiy usullardan biri bo'lib, ular ma'lumotlarni klassifikatsiya va regressiya uchun samarali tarzda ishlatiladi. Ular quyidagi asosiy jihatlar bilan ajralib turadi:

1. Tushunchasi va ishlash prinsipi

Qaror daraxtlari iyerarxik tuzilishga ega bo'lib, har bir tugun ma'lumotlarning muhim xususiyatlariga asoslangan qaror qabul qilish qoidalarini ifodalaydi. Daraxt quyidagicha tuziladi:

Ildiz tugun: Ma'lumotlarning barcha to'plamini o'z ichiga oladi va birinchi qoidaga asoslanib bo'linishni amalga oshiradi.

Ichki tugunlar: Ma'lumotlarni xususiyatlarga ko'ra keyingi kichikroq qismlarga ajratadi.

Barg tugunlar: Yakuniy qarorlarni yoki baholash natijalarini o'zida mujassam etadi.

2. Klassifikatsiya uchun qaror daraxtlari

Klassifikatsiya vazifalarida qaror daraxtlari ma'lumotlarni turkumlarga ajratadi. Masalan, kredit olish uchun ariza topshirgan mijozning layoqatli yoki nolayoqatligini aniqlashda foydalaniladi. Klassifikatsiyada quyidagi mezonlar qo'llaniladi:

Gini indeksi

Axborot darajasi (Entropy)

Axborot ortishi (Information Gain)

3. Regressiya uchun qaror daraxtlari

Regressiya vazifalarida qaror daraxtlari uzluksiz qiymatlarni bashorat qilish uchun ishlatiladi. Masalan, uy narxi yoki mahsulot sotuvlari hajmini prognozlash. Bo'linish mezoni sifatida:

MSE (Mean Squared Error)

MAE (Mean Absolute Error) kabi usullar qo'llaniladi.

4. Afzalliklari

Oddiy va intuitiv: Ma'lumotlarni ishlov berish va qarorlarni tushuntirish oson.

Kategorik va sonli ma'lumotlar bilan ishlash imkoniyati: Har xil turdagi ma'lumotlarni qayta ishlay oladi.

Moslashuvchanlik: Kichik hajmdagi ma'lumotlar bilan ham ishlashi mumkin.

Vizualizatsiya imkoniyati: Qaror daraxtlari vizual ko'rinishda berilib, tahlil qilishni osonlashtiradi.

5. Cheklovlari

Overfitting: Daraxt juda murakkab bo'lsa, o'rgatilgan ma'lumotlarga haddan tashqari moslashishi mumkin

Haddan tashqari oddiylashtirish: Daraxtni ortiqcha qirqish (pruning) modeli haddan tashqari oddiylashtirib yuborishi mumkin.

Katta ma'lumotlarda zaiflik: Katta hajmli yoki shovqinli ma'lumotlar bilan ishlashda ba'zan ansambl usullari (masalan, Random Forest yoki Gradient Boosting) samaraliroq bo'ladi.

Qaror daraxtlari nafaqat mustaqil usul sifatida, balki ansambl usullarining tarkibiy qismi sifatida ham foydalaniladi. Ular turli sohalarda, jumladan, tibbiyot, moliya, marketing va sanoatda qo'llaniladi.

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 layoqatligini 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.

Moliyaviy tahlil: Kredit risklarini baholash va firibgarlikni aniqlash.

Marketing: Mijozlarning xatti-harakatlarini segmentlash.

Ma'lumotlar tahlili: Murakkab tahlil va bashorat qilish jarayonlarini soddalashtirish.

FOYDALANILGAN DABIYOTLAR:

1. Onarkulov, M. K. (2024). INTEGRACIYA NEYRONNIX SETEY V RASPOZNAVANIY RECHI I RAZRABOTKA NOVIY METODOV. OBRABOVANIYE NAUKA I INNOVACIIONNYYE IDEYI V MIRE, 47(7), 193-197
2. Onarkulov, M. K., ugli Yusufov, M. A., & ugli Umirjonov, L. A. (2023). PRIMENENIYE NEYRONNIX SETEY V RASPOZNAVANIY RECHI. Educational Research in Universal Sciences, 2(3), 1206-1210.
3. Umarjon o'g, Y. Y. L. (2024). PROBLEMS AND MODERN TRENDS IN COMPUTING ENGINEERING. Multidisciplinary and Multidimensional Journal, 3(6), 17-21.
4. Onarkulov, M., & Nabijonov, A. (2024). DB2: KATTA HAJMDAGI MA'LUMOTLARNI SAMARALI BOSHQARISH. Инновационные исследования в науке, 3(5), 99-104.
5. Onarkulov, M., & G'oyibova, G. (2024). SQL SERVER ILOVASIDA UNIVERSAL MODELNI QO'LLASH VA KONVERTATSIYA ETISHNING SAMARALI YO'LLARI. Академические исследования в современной науке, 3(18), 147-152.
6. Karimberdiyevich, M. O. (2024). RBF TURLARINING O'QITISH ALGORITMI VA XOR MASALASI. worldly knowledge konferens, 8(1), 176-180.
7. Karimberdiyevich, M. O. (2024). GAUSS FUNKSIYASI. worldly knowledge konferens, 8(1), 239-244.
8. Karimberdiyevich, O. M. (2024). SQL TILIDA PROTSEDURA VA FUNKSIYALARDAN FOYDALANISHNING AHAMIYATI. worldly knowledge konferens, 8(1), 145-148.
9. Karimberdiyevich, O. M. (2024). BIR QATLAMLI NEYRON TO'RLARI VA ULARNI YARATISH USULLARI. IQRO INDEXING, 9(2), 104-108.
10. <https://www.mindonmap.com/uz/blog/what-is-decision-tree/>
11. <https://worldlyjournals.com/index.php/wkc/article/view/2852/4214>
12. <https://soff.uz/product/informatika-va-at-qarorlar-daraxti>