

DATA SCIENCE UCHUN ENG YAXSHI DASTURLASH TILLARI

Saidabonu Shokirova

Chirchiq Davlat Pedagogika Universiteti

Annotatsiya: *Ushbu maqola Data Science sohasida eng ko'p qo'llaniladigan dasturlash tillarini tahlil qiladi va ularning afzalliklari, kamchiliklari hamda qo'llanilish sohalari haqida ma'lumot beradi. Python, R, SQL, Julia, Java va Scala tillarining xususiyatlari ko'rib chiqilib, har bir tilning Data Science loyihalaridagi o'rni tahlil qilinadi. Maqola yangi boshlovchilar va tajribali mutaxassislar uchun qaysi tilni tanlash bo'yicha amaliy tavsiyalar beradi. Python'ning ko'p qirraliligi, R'ning statistik tahlildagi kuchi, SQL'ning ma'lumotlar bazasi bilan ishlashdagi muhimligi, Julia'ning yuqori tezligi, Java va Scala'ning Big Data loyihalaridagi afzalliklari alohida e'tiborga olinadi.*

Kalit so'zlar: *Data Science, dasturlash tillari, Python, R, SQL, Julia, Java, Scala, ma'lumotlar tahlili, mashinaviy o'rganish, statistik tahlil, Big Data, vizualizatsiya.*

KIRISH

Data Science — bu katta hajmdagi ma'lumotlarni tahlil qilish, ulardan foydali xulosalar chiqarish va sun'iy intellekt asosida innovatsion yechimlar ishlab chiqarishga xizmat qiluvchi zamonaviy soha. Bugungi kunda biznes, ilmiy tadqiqotlar va texnologik innovatsiyalar ma'lumotlarga asoslangan qarorlar qabul qilishga tayanadi[1]. Shu sababli, Data Science mutaxassislari uchun to'g'ri dasturlash tilini tanlash loyihalarni samarali amalga oshirishda muhim ahamiyatga ega. Har bir dasturlash tili o'ziga xos afzalliklarga ega bo'lib, ma'lum vazifalar uchun boshqalarga nisbatan ustunlik qiladi. Ushbu maqola Data Science sohasida eng ko'p qo'llaniladigan Python, R, SQL, Julia, Java va Scala tillarini tahlil qiladi, ularning kuchli va zaif tomonlarini ko'rib chiqadi hamda qaysi tilni qachon tanlash bo'yicha amaliy yo'l-yo'riq beradi[2].

Data Science uchun dasturlash tillarining taqqosiy tahlili

Python: Ko'p qirralilik va kuchli ekotizim

Python hozirgi kunda Data Science sohasida eng ko'p ishlatiladigan dasturlash tillaridan biridir. Uning sintaksisi sodda va o'rganish osonligi uni yangi boshlovchilar uchun ideal vositaga aylantiradi. Python'ning NumPy, Pandas, Matplotlib, Scikit-learn, TensorFlow va PyTorch kabi boy kutubxonalari orqali ma'lumotlarni qayta ishlash, vizualizatsiya qilish va mashinaviy o'rganish modellarini yaratish imkoniyati mavjud. Shu bilan birga, Python umumiy maqsadli dasturlash tili bo'lib, Data Engineering va Web development kabi boshqa yo'nalishlar uchun ham keng qo'llaniladi[3]. Ammo, Pythonning bajarilish tezligi C/C++ yoki Julia kabi tillarga nisbatan pastroq bo'lishi mumkin.

R: Statistika va vizualizatsiyaga ixtisoslashgan

R dasturlash tili ayniqsa statistik tahlil va ilmiy tadqiqotlar uchun mo'ljallangan. Bu tilda ma'lumotlarni tahlil qilish, regressiya modellarini qurish, klasterlash va boshqa

statistik usullarni bajarish juda qulay. ggplot2, dplyr, caret kabi kuchli paketlar orqali R statistik vizualizatsiyalarda yetakchi hisoblanadi. R'ning kamchiligi sifatida esa uning umumiy maqsadlar uchun mos emasligi va katta hajmdagi ma'lumotlar bilan ishlashda sekin ishlashi keltirilishi mumkin.

SQL: Ma'lumotlar bazasi bilan samarali ishlash

SQL (Structured Query Language) – ma'lumotlar bazasini boshqarish va ulardan ma'lumotlarni samarali tarzda chiqarib olish uchun standart til hisoblanadi. Data Science loyihalarining aksariyatida tahlil qilinadigan ma'lumotlar avvaldan ma'lumotlar bazasida saqlanadi. Shu bois, SQL bilimiga ega bo'lish data analiz jarayonida zaruratga aylangan. Biroq, SQLning imkoniyatlari faqat ma'lumotlarni olish va ularni transformatsiya qilish bilan cheklangan bo'lib, model qurish yoki mashinaviy o'rganish jarayonlarida u bevosita qo'llanilmaydi.

Julia: Yuqori samaradorlik va ilmiy hisob-kitoblar uchun moslik

Julia – nisbatan yangi dasturlash tili bo'lib, ayniqsa ilmiy hisob-kitoblar va yuqori samaradorlik talab qilinadigan Data Science vazifalari uchun ishlab chiqilgan. Julia kompilyatsiyalangan til bo'lib, Python va R'ga qaraganda ancha tez ishlaydi. U matematik formulalarni sodda ifodalash imkonini beradi va parallel hisoblashda yaxshi ko'rsatkichlarga ega[4]. Ammo Julia'ning kutubxonalar ekotizimi hali rivojlanayotgan bosqichda bo'lib, ko'plab tayyor resurslar hali yetarli emas.

Java va Scala: Big Data va korporativ ilovalar uchun kuchli vosita

Java va Scala tillari, ayniqsa, Big Data tahliliga ixtisoslashgan. Apache Hadoop va Apache Spark kabi texnologiyalar Java yoki Scala asosida ishlaydi. Scala, Java'ga nisbatan ixcham sintaksisga ega bo'lsa-da, u bilan moslashishda dasturchidan funksional dasturlash konseptlarini tushunishni talab qiladi. Bu tillar ayniqsa katta hajmdagi real vaqtli ma'lumotlarni qayta ishlash, tarmoq tizimlari bilan ishlash va katta miqyosdagi korporativ tizimlarda Data Science loyihalarini amalga oshirish uchun qo'llaniladi.

Data Science sohasida ishlatiladigan har bir dasturlash tili ma'lum kontekstda o'z ustunligiga ega. Quyidagi holatlar tilda tanlov qilishda foydali yo'l-yo'riq bo'la oladi:

- Python — agar siz yangi boshlovchi bo'lsangiz, mashinaviy o'rganish, chuqur o'rganish (deep learning), data engineering yoki avtomatlashtirish sohalariga qiziqsangiz, Python — bu eng maqbul tanlov. Ko'plab onlayn kurslar va jamoa yordami mavjud.

- R — siz statistik modellash, ma'lumotlar vizualizatsiyasi yoki akademik tadqiqotlar bilan shug'ullansangiz, R sizga yuqori aniqlikdagi tahliliy natijalarni taqdim etadi. Tibbiyot, ijtimoiy fanlar va bioinformatika sohalarida keng qo'llaniladi.

- SQL — ma'lumotlarni olish, transformatsiya qilish (ETL), BI (Business Intelligence) tizimlarida ishlashda asosiy vosita bo'lib xizmat qiladi. Har qanday Data Analyst yoki Data Scientist mutaxassisi uchun SQL bilimlari zarurdir.

- Julia — siz yuqori samaradorlik talab qilinadigan ilmiy hisob-kitoblar bilan shug'ullanayotgan bo'lsangiz, Julia katta hajmdagi ma'lumotlarni tez va samarali qayta ishlash imkonini beradi. Hisoblash fizikasi, kvant mexanikasi, moliyaviy modellash tirishda foydali.

•Java yoki Scala — agar siz Big Data platformalarida (Apache Hadoop, Apache Spark) yoki korporativ darajadagi real vaqtli tahlil tizimlarida ishlayotgan bo'lsangiz, bu tillar samarali va barqaror yechimlar yaratishda foyda keltiradi.

Data Science sohasida muvaffaqiyatga erishish uchun nafaqat biror bitta dasturlash tilini puxta egallash, balki turli tillarning kuchli jihatlarini tushunish va ularni kerakli holatda qo'llay olish ko'nikmasi muhimdir. Quyidagi tavsiyalar ushbu yo'nalishda o'z bilimini chuqurlashtirmoqchi bo'lganlar uchun foydali bo'lishi mumkin[5]:

- Yangi boshlovchilar dastlab Python yoki R tilidan boshlashi maqsadga muvofiq.
- Ilmiy yoki statistik tahlil qilishni maqsad qilganlar uchun R, Julia eng mos keladi.
- Katta hajmdagi sanoat ma'lumotlari bilan ishlashni istaganlar Java yoki Scala texnologiyalarini o'zlashtirishi zarur.

•Real loyihalarda ishlayotgan mutaxassislar esa tillararo integratsiyani o'rganib, turli texnologiyalarni birgalikda ishlata olish ko'nikmasini rivojlantirishlari lozim.

XULOSA. Data Science sohasida to'g'ri dasturlash tilini tanlash loyiha muvaffaqiyati uchun muhim ahamiyatga ega. Python o'zining ko'p qirraliligi va boy kutubxonalari bilan yangi boshlovchilar va tajribali mutaxassislar uchun eng maqbul tanlovdir. R statistik tahlil va vizualizatsiyada, SQL ma'lumotlar bazasi bilan ishlashda, Julia yuqori samaradorlik talab qilinadigan ilmiy hisob-kitoblarda, Java va Scala esa Big Data va korporativ loyihalarda muhim o'rin tutadi. Har bir til o'ziga xos afzalliklarga ega bo'lib, ma'lum vazifalar uchun boshqalarga nisbatan ustunlik qiladi. Data Science mutaxassisi sifatida muvaffaqiyatga erishish uchun bir nechta tillarning kuchli tomonlarini o'rganib, ularni loyiha talablariga mos ravishda qo'llay olish zarur. Yangi boshlovchilar Python yoki R tilidan boshlashlari, tajribali mutaxassislar esa tillararo integratsiya va maxsus sohalarga ixtisoslashgan tillarni o'zlashtirishlari tavsiya etiladi. Shu yo'l bilan Data Science sohasida samarali va innovatsion yechimlarni ishlab chiqish mumkin.

FOYDALANILGAN ADABIYOTLAR:

1. Matthes, E. (2019). Python Crash Course: A Hands-On, Project-Based Introduction to Programming (2nd ed.). No Starch Press.
2. Guttag, J. V. (2016). Introduction to Computation and Programming Using Python (2nd ed.). MIT Press.
3. D. P. Kroese, T. Taimre, and Z. I. Botev. Handbook of Monte Carlo Methods. John Wiley & Sons, New York, 2011.
4. R. G. Bartle. The Elements of Integration and Lebesgue Measure. John Wiley & Sons, Hoboken, 1995.
5. Raqamli Avlod. (n.d.). Data Science sohasiga qiziqqanlar uchun yo'l xaritasi. Retrieved from <https://raqamliavlod.uz>
6. R. Y. Rubinstein. The cross-entropy method for combinatorial and continuous optimization. Methodology and Computing in Applied Probability, 2:127–190, 1999.

7. Mohirdev. (2024). Scala va Spark: Data injenerlar uchun kurslar. Retrieved from <https://mohirdev.uz>
8. Pyblog. (2022). 2022-yilda mashhur dasturlash tillari: TOP-6 talik. Retrieved from <https://pyblog.uz>