

4-lesson. Report

Ushbu darsda Machine Learning loyihasini noldan boshlab yakuniy natijagacha olib borish jarayoni, ya'ni ML project ketma-ketligiga Data Preprocessingbosqichi keng yoritildi. Shuningdek, ML loyihalarda ishlatiladigan asosiy Python kutubxonalar bilan tanishildi.

Machine Learning loyihasi bir nechta muhim bosqichlardan iborat bo'ladi. Ushbu bosqichlar ketma-ket bajarilishi modelning sifatli natija berishini ta'minlaydi.

1. Data Collection (Ma'lumot to'plash)

Ma'lumotlarni turli manbalardan olish mumkin:

- So'rovnomalari orqali
- Web saytlar (web scraping)
- Tayyor datasetlar (Kaggle, GitHub)
- Mijoz tomonidan taqdim etilgan ma'lumotlar

Ma'lumotlar bilan ishlashdan oldin fayllar bir xil joyda saqlanadi va Pandaskutubxonasi orqali yuklanadi.

2. Data bilan tanishish

Bu bosqichda:

- Dataset tuzilishi o'r ganiladi
- head()va tail()funksiyalari orqali ma'lumotlar ko'rib chiqiladi
- Keraksiz ustunlar aniqlanib olib tashlanadi

3. Data Preprocessing (Ma'lumotlarni tayyorlash)

Data Preprocessing — model ishlashi uchun eng muhim bosqichlardan biridir.

Missing values (tushirib qoldirilgan qiymatlar) bilan ishlash:

- Mean— o'rtacha arifmetik qiymat bilan to'ldirish (raqamli ustunlar uchun)
- Mode— eng ko'p takrorlangan qiymat bilan to'ldirish
- Median— o'rtadagi qiymat bilan to'ldirish (raqamli ma'lumotlar uchun)
- Fixed value— foydalanuvchi belgilagan qiymat bilan to'ldirish
- Drop— yo'qolgan qiymatlarni o'z ichiga olgan qator yoki ustunlarni o'chirish

4. Model Selection va Training

Bu bosqichlarda:

- Mos ML modeli tanlanadi
- Ma'lumotlarning 70% qismi training uchun ishlatiladi
- Model o'qitiladi

5. Prediction, Evaluation va Testing

- Prediction— model orqali bashorat qilish
- Evaluation— model sifatini baholash (15%)
- Testing— modelni test qilish (15%)

6. Deployment va Monitoring

Tayyor model real tizimga joyланади (deployment) va uning ishlashi doimiy kuzatib boriladi (monitoring).

Python kutubxonalari (libraries): Pandas - ma'lumotlarni yuklash, tozalash, va tahlil qilish, NumPy - matematik va statistik va statistik hisob-kitoblar, matplotlib - grafik va vizualizatsiyalar, Seaborn - statistik ma'lumotlarni vizual ko'rsatish, Scikit-learn - ML modellarini yaratish va o'qitish

Ma'lumotlar quyidagi xususiyatlarga ega bo'lishi kerak: Structured (tuzilgan), to'liq, katta soha bilan bog'liq, ishonchli

Ushbu dars davomida: ML project ketma-ketligini, Data preprocessing tushunchasini, Missing values bilan ishlashning 5 xil usulini, asosiy Python ML kutubxonalarini, dataset bilan ishlash asoslarini o'rgandik.