MACHINE LEARNING EXAM 1

Tursunov Akhmadjon

1 Model Reporting

[utf8 inputenc

* Ushbu bo'lim loyihaning qisqacha tavsifi, maqsadi va asosiy natijalarini taqdim etadi.
* Maqsad:
* Ushbu loyiha maqsadi parametrlarga, asoslanib avtomobillarning US(Amerikada) da sotilish yoki sotilmasligini aniqlash(klassifikatsiya qilish)

Ma'lumotlar Haqida Umumiy Ma'lumot

* Bu bo'limda ma'lumotlar to'plami, uning tuzilishi va modelda ishlatilgan xususiyatlar haqida umumiy ma'lumot beriladi.
* Ishlatilgan Ma'lumotlar To'plami:
* Ma'lumotlar survey orqali individual shaxslardan olingan bo'lib, ularning yoshi, oyligi, yashash hududi vaboshqa parametrlarga asoslangan
* Xususiyatlar:
* Sales, ComPrice, Income, Advertising, Population, Price, Shelveloc, Age, Education, Urban, US
* Asosiy Xususiyatlar:
* Income: Individual oylik daromadi
* Price: Avto narxi
* Urban: Yashash hududi rivojlangan darajasi
* Ma'lumotlar Tahlili:
* Kategorik o'zgaruvchilar (masalan, Price, ShelveLoc, Education, Urban, US) uchun Label kodlash amalga oshirildi.
* Malulotlarda yetishmagan qiymatlar (ShelveLoc va US) aniqlandi va ularni mode. Qiymati bilan to’ldirildi

Modelni Rivojlantirish

* Bu bo'limda tanlangan mashinalarni o'qitish modelini, trening jarayonini va baholashni tavsiflashadi.
* Tanlangan Model: Logistic Regression, Decision Tree, KNN, Random Forest Modellari ishlatildi
* Tanlash Sababi:
* Logistic Regression modeli kichik va o‘rta o‘lchamdagi ma’lumotlarda juda tez ishlaydi
* Decision Tree modeli, tushunishga oson va ma’lumot ichida qaysi xususiyatlar eng muhimligini avtomatik aniqlaydi
* KNN modeli xususiyatlar orasidagi bog‘liqlikni o‘z-o‘zidan aniqlash xususiyatiga ega
* Random Forest katta hajmdagi va turli xil ma’lumotlar bilan yaxshi ishlaydi, haddan tashqari moslashuvchanlikni oldini oladi.
* Trening:
* Model ma'humotlar to'plamining 70%, 15%, 15% ga bo'lingan holda o'qitildi.
* Modelni optimallashtirish uchun gidroparameterlarni sozlash amalga oshirildi.

Modelni Baholash

* Accuracy Score: [Qiymat].

Xulosa

* Modelning umumiy muvaffaqiyatini va kelajakdagi takomillashtirish imkoniyatlarini umumlashtirish.
* Xulosa:

Logistic Regression aniqlik darajasi 90%

Decision Tree ning aniqlik darajasi 87% edi, ammo Hyper parametr tuningdan so’ng uning ham ko’rsatgichi 90% yetdi

KNN aniqlik darajasi 67% ga teng, scaling olib borilganiga qaramay eng past ko’rsatgichga ega.

Random forest ning aniqlik darajasi 90% ga teng.