MACHINE LEARNING EXAM 1

Tursunov Akhmadjon

1 Model Reporting

* Ushbu bo'lim loyihaning qisqacha tavsifi, maqsadi va asosiy natijalarini taqdim etadi.
* Maqsad:
* Ushbu loyiha maqsadi parametrlarga, asoslanib taxi xizmatining narxini bashorat(predict) qilish dir

Ma'lumotlar Haqida Umumiy Ma'lumot

* Bu bo'limda ma'lumotlar to'plami, uning tuzilishi va modelda ishlatilgan xususiyatlar haqida umumiy ma'lumot beriladi.
* Xususiyatlar:
* Trip\_Distance\_km, Time\_of\_Day, Day\_of\_Week, Passenger\_Count, Traffic\_Conditions, Weather, Base\_Fare, Per\_Km\_Rate, Per\_Minute\_Rate, Trip\_Duration\_Minutes, Trip\_Price
* Asosiy Xususiyatlar:
* Trip\_Distance\_km: Yetib borish kerak bo’lgan manzilgacha masofa
* Passenger\_Count: Mijozlar soni
* Trip\_Duration\_Minutes: Sayohat uchun ketgan vaqt
* Trip\_Price: Narx
* Ma'lumotlar Tahlili:
* Kategorik o'zgaruvchilar (Trip\_Distance\_km, Time\_of\_Day, Day\_of\_Week, Passenger\_Count, Traffic\_Conditions, Weather, Trip\_Price).
* Passenger\_Count column ni alohida o’rganib chiqildi va undagi “four” va “two” so’zlari ‘4’, ‘2’ ga almashtirildi va yetismaydigan ma’lumotlor “mode()” funksiyasi bilan to’ldirildi. Ohirida Column ‘data type’ni object dan int ga o’tkzildi.
* Trip\_Distance\_km, Trip\_Price columnlarida ($, km) elementlar qatnashganligi sababli float -> object bo’lib korinmoqda. Muammon yechish uchun har bir qatorni tekshirib chiqib sonlardan boshqa hama elementlarni “” bo’sh joy bilan almashtirib chiqamiz, so’ng yetishagan qiymatlarni (mean()) funsiyasi bilan to’ldiramiz.
* Time\_of\_Day, Day\_of\_Week, Traffic\_Conditions, Weather ‘column’lar object va ularni high yoki low cardinality ekanligini tekshirib, low cardinality ekanligi sababli One-hot encodingdan foydalanamiz.

Modelni Rivojlantirish

* Bu bo'limda tanlangan mashinalarni o'qitish modelini, trening jarayonini va baholashni tavsiflashadi.
* Tanlangan Model: Decision Tree Modelini ishlatildi
* Tanlash Sababi:
* Decision Tree modeli, tushunishga oson va ma’lumot ichida qaysi xususiyatlar eng muhimligini avtomatik aniqlaydi
* Trening:
* Model ma'humotlar to'plamining 70%, 15%, 15% ga bo'lingan holda o'qitildi.
* Modelni optimallashtirish uchun gidroparameterlarni sozlash amalga oshirildi.

Modelni Baholash

* Model train qilingandan so’ng uni tekshirdik va quyidagi xatolik natijalarini oldik

mse:312.6512200536612

r2:0.8445700437634959

* Model xatoligini tushirish uchun data setni sclaning qildik va so’ng uni tekshirdik va quyidagi xatolik natijalarini oldik

mse:302.56305594539447

r2:0.8495852261945303

Xulosa

* Xulosa:

|  |  |  |
| --- | --- | --- |
|  | Standard accuracy | Accuracy after scaling |
| MSE | 312.65 | 302.56 |
| R2 | 0.844 | 0.849 |

Hatolik katta, modelni yaxshilash uchun HyperParamet tuning, feature engineering dan foydalanishiz, yoki boshqa model tanlashimiz kerak.