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 ‘Big Market’dan olingan ma’lumotlarga asoslanib, Item outlet sales (Narxini) bashorat qilish.

**Features Haqida Umumiy Ma'lumot**

Ma'lumotlar to'plami 8523 tani tashkil etib, 10 ta xususiyat(features)dan va Target feature “**Item\_Outlet\_Sales**” iborat.

* Raqamli o'zgaruvchilar: Item\_Weight, Item\_Visibility, Item\_MRP, Outlet\_Establishment\_Year, Item\_Outlet\_Sales.
* Kategorik o'zgaruvchilar**:** Item\_Identifier, Item\_Fat\_Content, Item\_Type, Outlet\_Identifier, Outlet\_Size, Outlet\_Location\_Type, Outlet\_Type.

|  |  |
| --- | --- |
|  |  |

Feature.lar o’rtasidagi bo’g’liqlikni ko’rishimiz mumkinki u Chiziqli emas. Shu sababli biz Non-Linear Modellardan foydalanamzi

**Data Preprocessing jarayoni (Decision Tree)**

* Missing value bilan ishlash:

**Item\_Type va Outlet\_Size** featureimizdagai Null valuelarni mode() orqali, **Item\_Weight f**eatureimizdagai Null valuelarni mean() orqali to’ldirdik

* Coding:

‘Outlet\_Size', 'Outlet\_Location\_Type', 'Outlet\_Type’ feature.larni One-hot encoding (cordinalaty<4) bilan qoganlarini label encoding bilan raqamli ko’rinishga o’tkazdik. Bu data setimizni toliqligicha numerical o’rinishga o’tkazishga yordam beradi.

* Scaling:

Malumotlarimizni Standard scaling orqali muvozanatga olib keldik.

* Maʼlumotlarni boʻlish:

Dataset imizni 70% ini traingn uchun qolgan 30% izini test uchun bo’lib oldik

**Data Preprocessing jarayoni (LGBoost)**

* Coding:

Barcha ‘object’ type dagi feature.larimizni ‘category’ga o’tkazdik

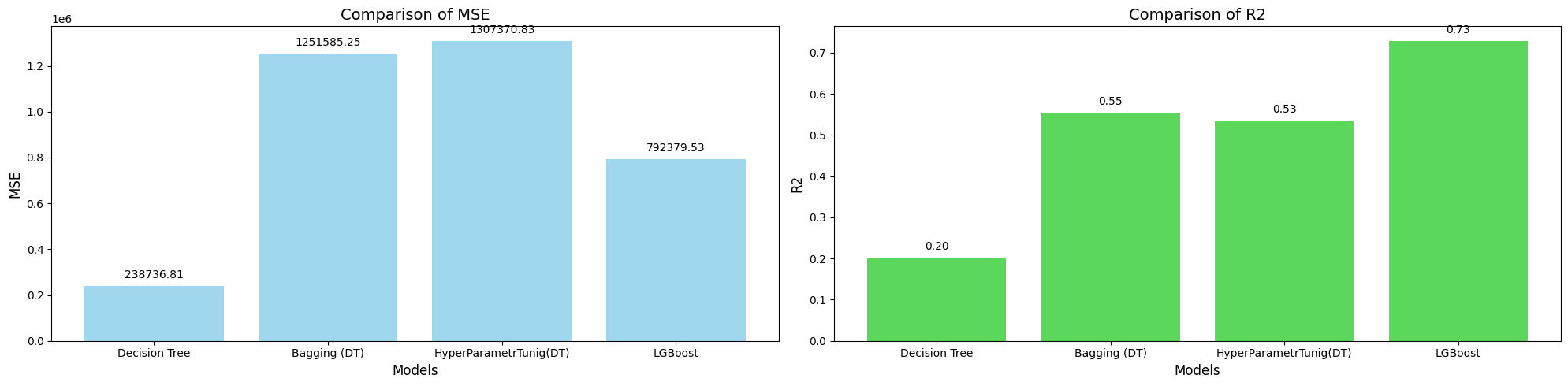
**Modelni Rivojlantirish**

* Tayyor data set imizni Decision Tree Regression, LGBoost modellarida o’qitib, ularni o’zaro solishtiramiz.
* LGBoost da data preprocessingjarayoni ancha kam bo’ldi va ularning qulay tomoni hisoblanadi. Undan tashqari LGBoost teskor va kichik data setlar uchun maqul model hisoblanadi
* Decision tree daraxt strukturasiga asoslanib ma’lumotlarni juda yaxshi o’rganadi.

Decision Tree da model o’qitilgandan so’ng mse=238736.8131542644, r2=20% natija berdi, Bagging orqali modelni yaxshilashga urinib ko’rdik va mse=1251585.2461680584, r2=55% ga yaxshilandi, ammo bu yetarli emas.Hyper parametr tuning dan so’ng natijamiz ko’tarildi ammo sezilarmi emas.

Modelni Baholash

* Mean square error
* R2 score

Xulosa

Natijadan ko’rinib turibdiki LGBoost biza uchun ma’qul model.