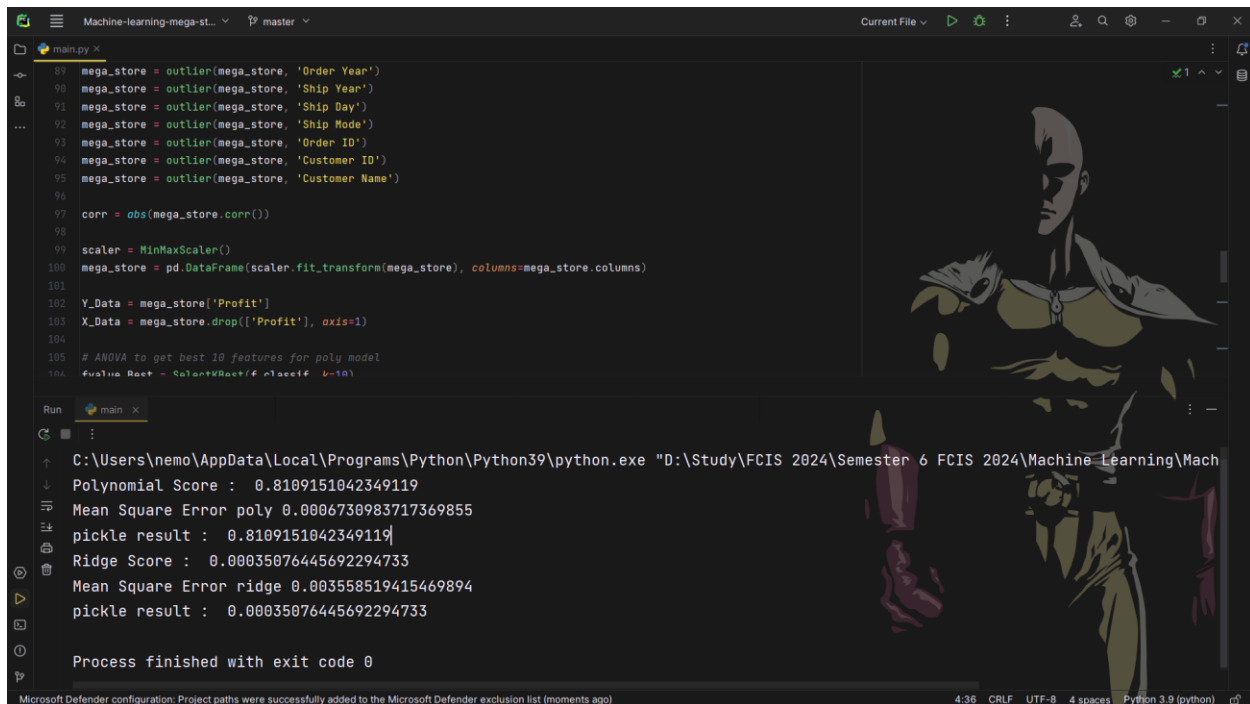


20201700994 يوسف احمد بسيم علي غيط  
20201700503 علي سامح سعد عبدالرحمن  
20201700463 عبدالله ماجد محمد عوض  
20201700467 عبدالمجيد حاتم عبدالمجيد عبدربه  
20201701014 يوسف ضياء الدين محمد حسين  
20201700500 علي ابراهيم علي ابراهيم

- 1) Drop date column and add (day ,month ,year) columns in DateFormat Function ,three columns better than one column (date).
- 2) Drop outliers greater than 75% and lower than 25% with outlier Method
- 3) Check the how many countries in country column ,if the same drop the column
- 4) Choose the values of the CategoryTree to (MainCategory ,Subcategory) , then drop the original column
- 5) Encoding all columns with string values
- 6) Applying DateFormat function on ShipDate and OrderDate
- 7) Dropping 5 columns with lowest absolute correlation (Shipmonth ,Ordermonth ,Orderday ,Rowid ,Productname) with the Profit Column.
- 8) Scaling all Data using MinMax Method
- 9) Splitting Data 20% Test - 80% Training

- 10) Feature Selection using Anova model with the highest 10 for the polynomial model
- 11) Generating Polynomial Model with Degree 3 to avoid overfitting.
- 12) Generating Ridge Model with the worst 8 features (ship\_mode ,ship\_day ,ship\_year ,order\_year ,customer\_id ,customer\_name ,order\_id ,postal\_code) from the Correlation table.
- 13) Saving each model using Pickle library
- 14) Polynomial model is better than ridge because of the feature ,then better accuracy.



```
main.py x
89 mega_store = outlier(mega_store, 'Order Year')
90 mega_store = outlier(mega_store, 'Ship Year')
91 mega_store = outlier(mega_store, 'Ship Day')
92 mega_store = outlier(mega_store, 'Ship Mode')
93 mega_store = outlier(mega_store, 'Order ID')
94 mega_store = outlier(mega_store, 'Customer ID')
95 mega_store = outlier(mega_store, 'Customer Name')
96
97 corr = abs(mega_store.corr())
98
99 scaler = MinMaxScaler()
100 mega_store = pd.DataFrame(scaler.fit_transform(mega_store), columns=mega_store.columns)
101
102 Y_Data = mega_store['Profit']
103 X_Data = mega_store.drop(['Profit'], axis=1)
104
105 # ANOVA to get best 10 features for poly model
106 f_valua_Rest = SelectKBest(f_olseoff k=10)

Run
C:\Users\nemo\AppData\Local\Programs\Python\Python39\python.exe "D:\Study\FCIS 2024\Semester 6 FCIS 2024\Machine Learning\Mach
Polynomial Score : 0.8109151042349119
Mean Square Error poly 0.0006730983717369855
pickle result : 0.8109151042349119
Ridge Score : 0.0035076445692294733
Mean Square Error ridge 0.003558519415469894
pickle result : 0.0035076445692294733

Process finished with exit code 0

Microsoft Defender configuration: Project paths were successfully added to the Microsoft Defender exclusion list (moments ago)
4:36 CRLF UTF-8 4 spaces Python 3.9 (python)
```