JEDDAH UNIVERSITY - JEDDAH

College of Computer Science and Engineering, Information Systems Department

Lab _#2

Applied Machine Learning **CCDS 322**

Rubric of Lab2

Name: Joud Mohammed Alahmari ID: 2008071	**Max Marks**	**Obtained Marks**	Comments
CONTENTS: Question 1			
Question 1(1)	2		
Question 1(2)	6		
Question 1(3)	3		
Question 1(4)	2		
Question 1(5)	2		
Question 1(6)	2		
Question 1(7)	2		
TOTAL MARKS	**19**		
ORGANISATION			
• PDF or DOCX file and source code file, Name your file as <Assignment1_ID>, cover page body, headings, fonts, style, exhibits, length, page numbers, etc.	2		

• Informative Screen shots, tables, figures, etc..	2		
TOTAL MARKS	**23**		
PENALTY			
• Similarity index • Late submission			
FINAL MARKS	**15**		

Answer all questions:

Document Requirements :

- ❖ The Lab is submitted electronically through Blackboard. Each late assignment will be penalized 1 point per late working day.
- ❖ The format of document:
  - ✓ Softcopy of document as DOCX file and source code file. Name your file as < Lab5 _ID>. send me it by using blackboard
  - ✓ Cover page with the members' names, ID
  - ✓ Text times roman font 12 or equivalent
  - ✓ Line spacing 1.5 and must include page numbering
  - ✓ Properly bound (Do not submit loose pages in folders. The instructor will not be responsible if any of the pages are missing).
  - ✓ Indent the first line of all paragraphs
  - ✓ Justify all paragraphs

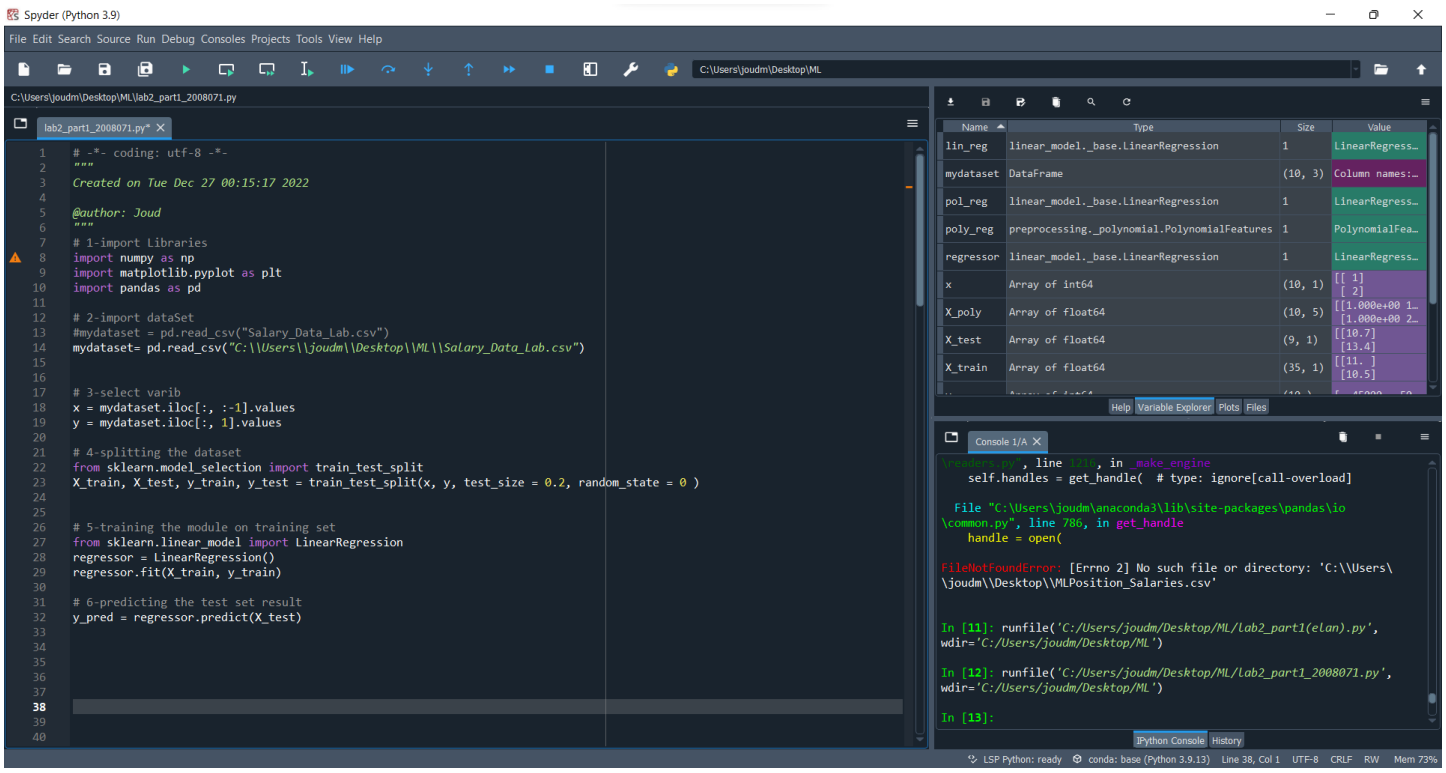
Question 1

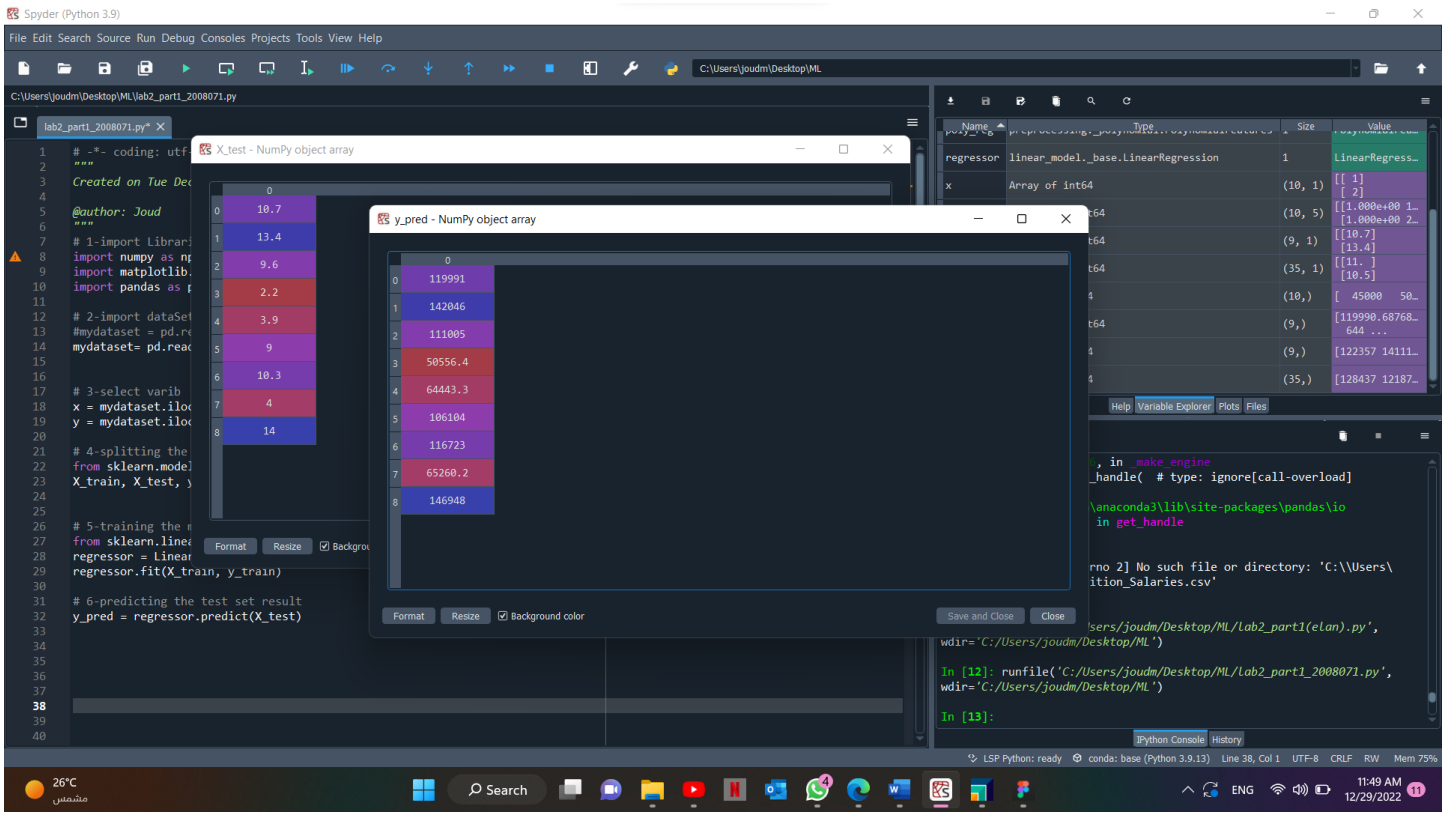
You are working in a specific company as a data analyst. This company has an available position for country manager and new person applied for this position. He/She has been working for 3 years as a country manager in another company and his/her salary is 133,000$. HR sent you the positions' salary dataset and ask you if he/she deserve the 133,000$? Consider his\her level on 5.5.

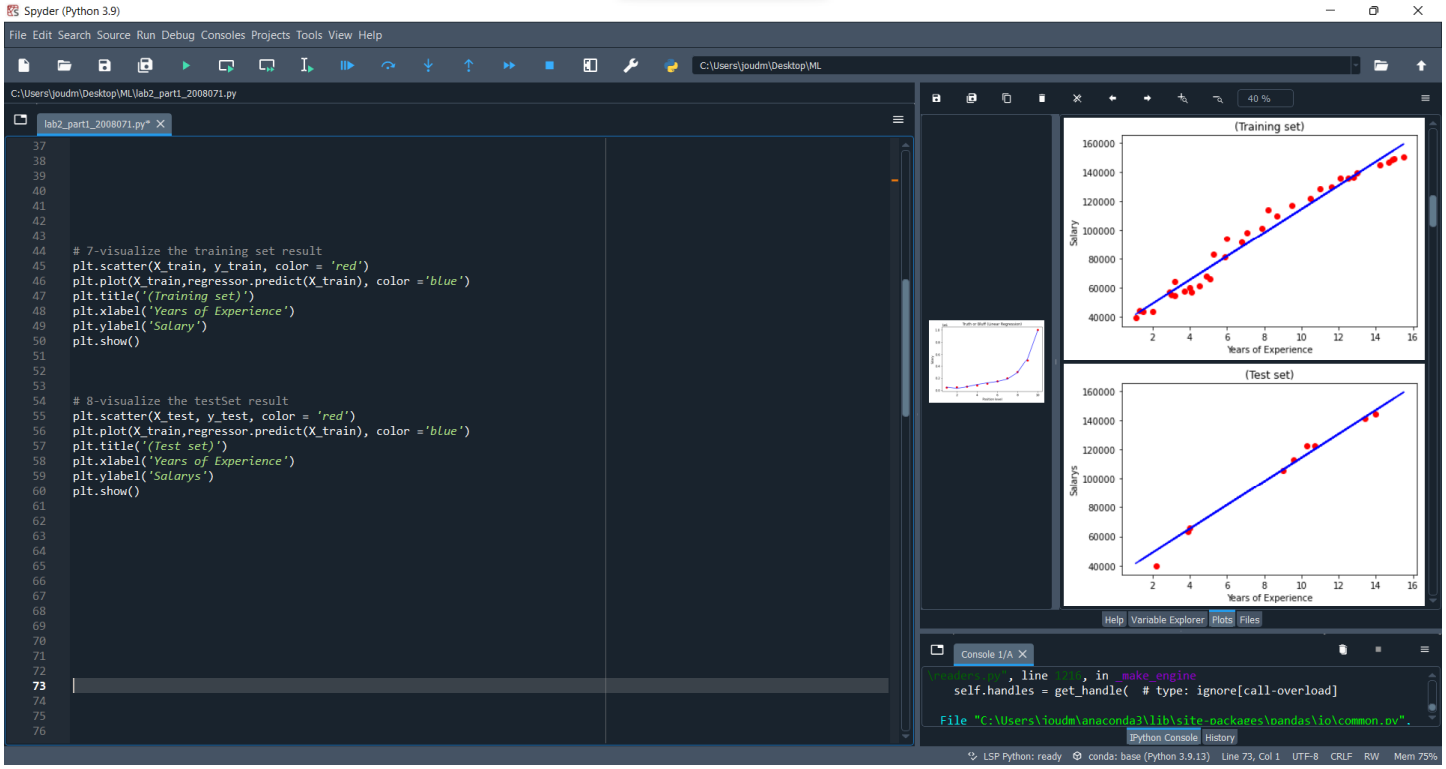
1. Download the attached dataset.

2. Predict the new country manager's salary using the following:

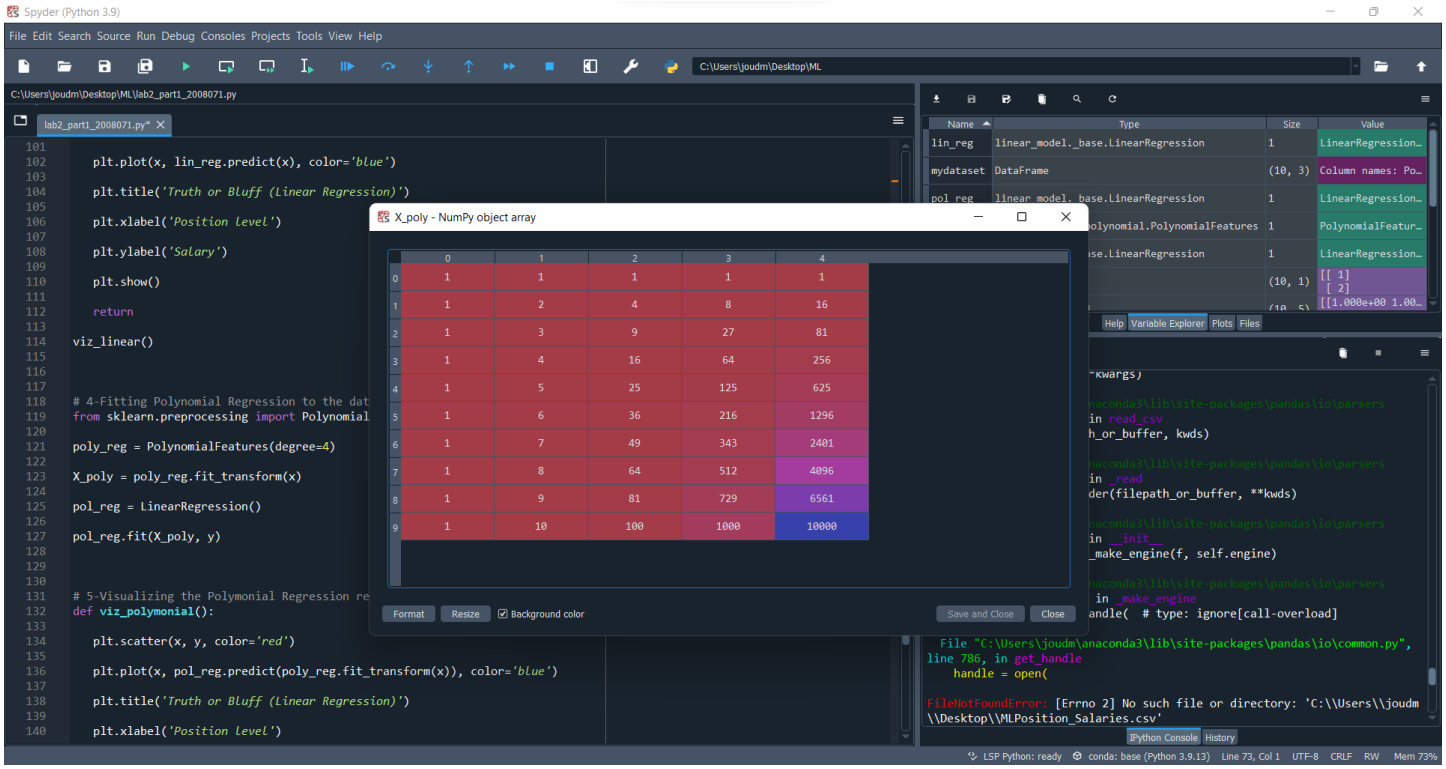
   a. Simple linear regression





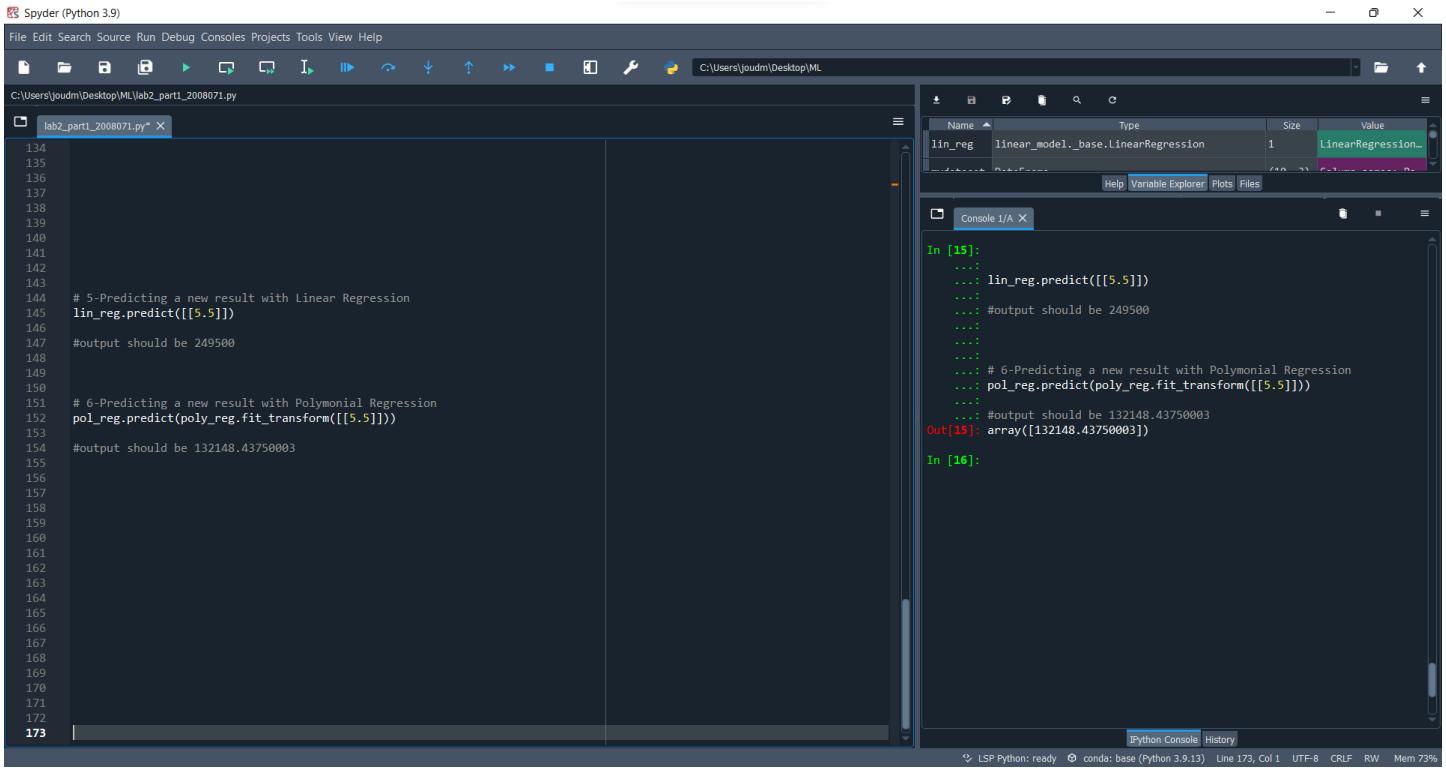


b. Polynomial linear regression.



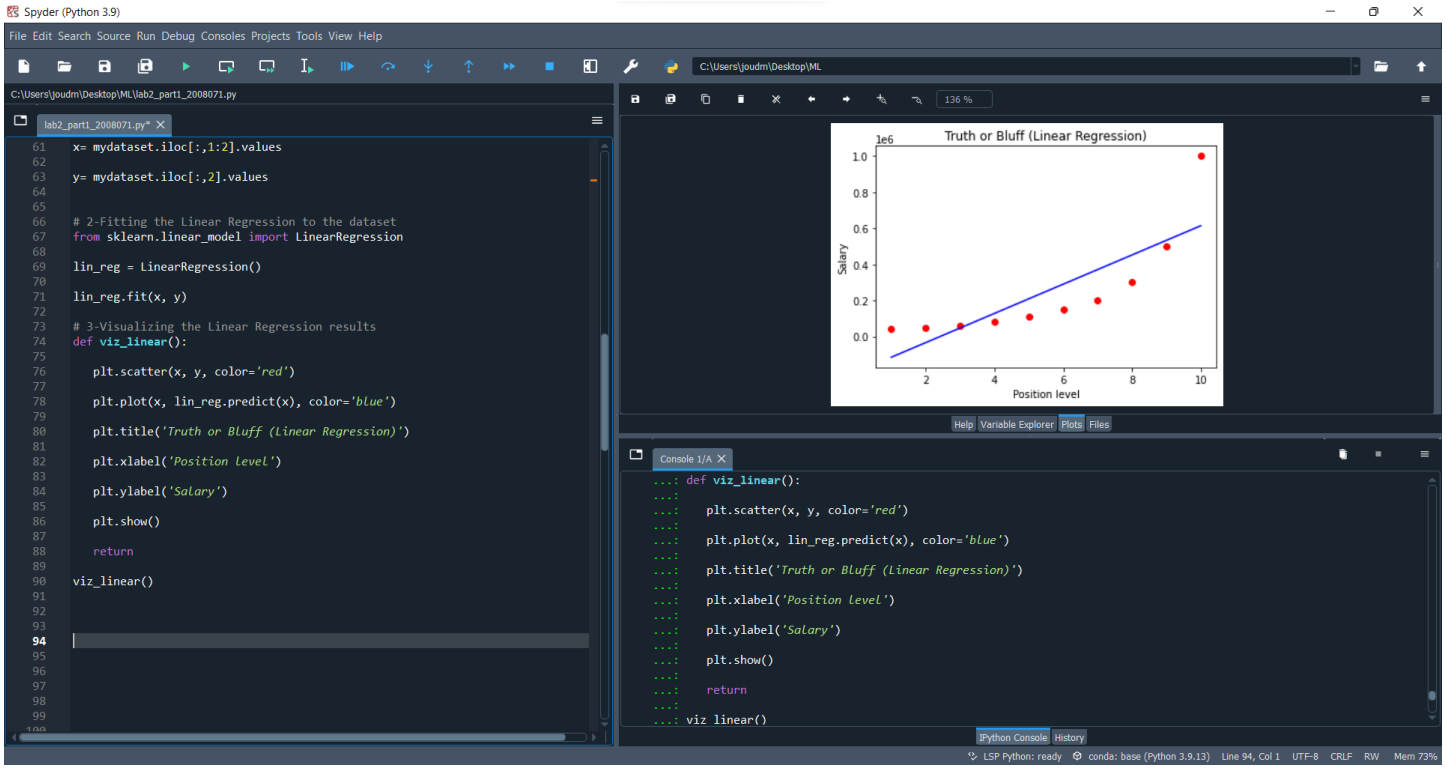
c. Support vector regression.

3. Write or show screenshot of the prediction results.

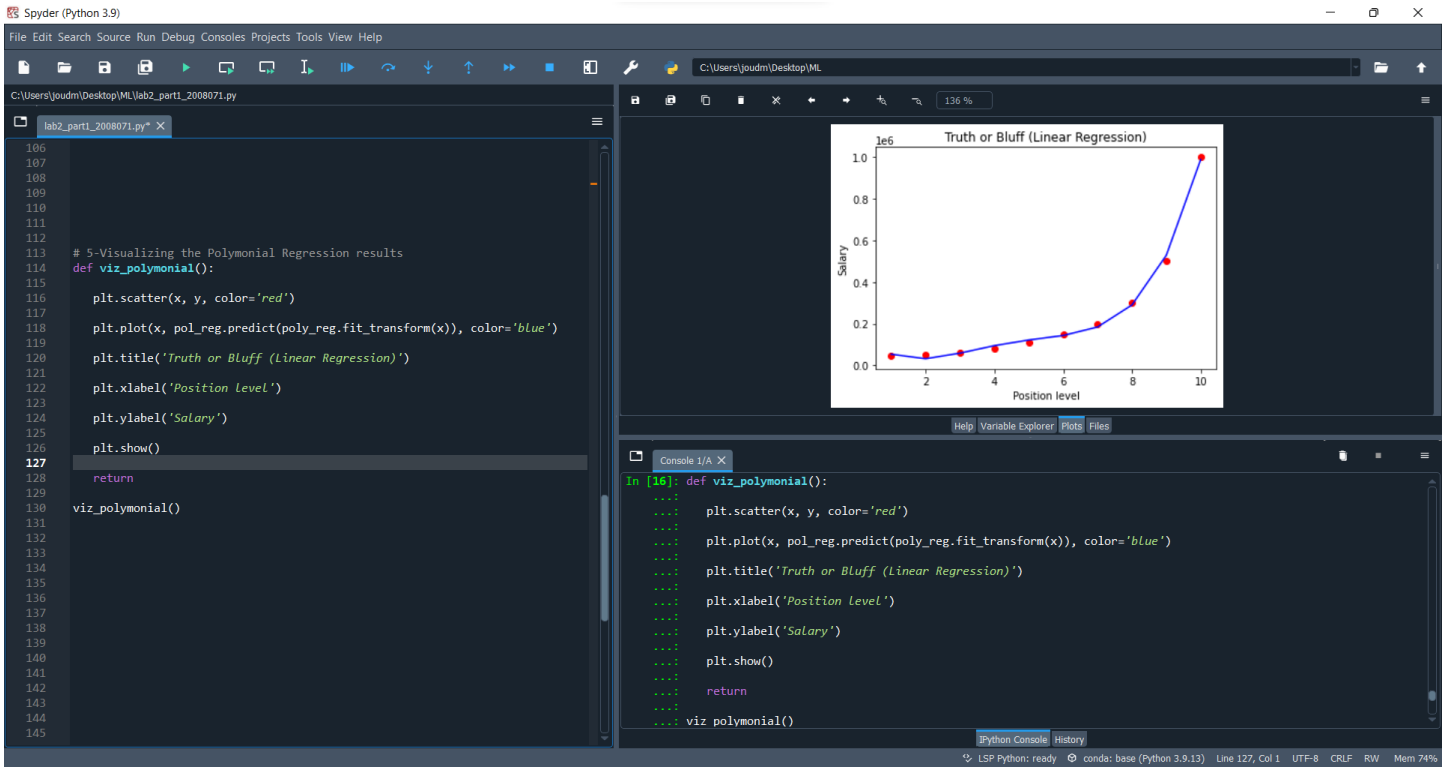


4. Visualize the simple linear regression result (show screenshot).

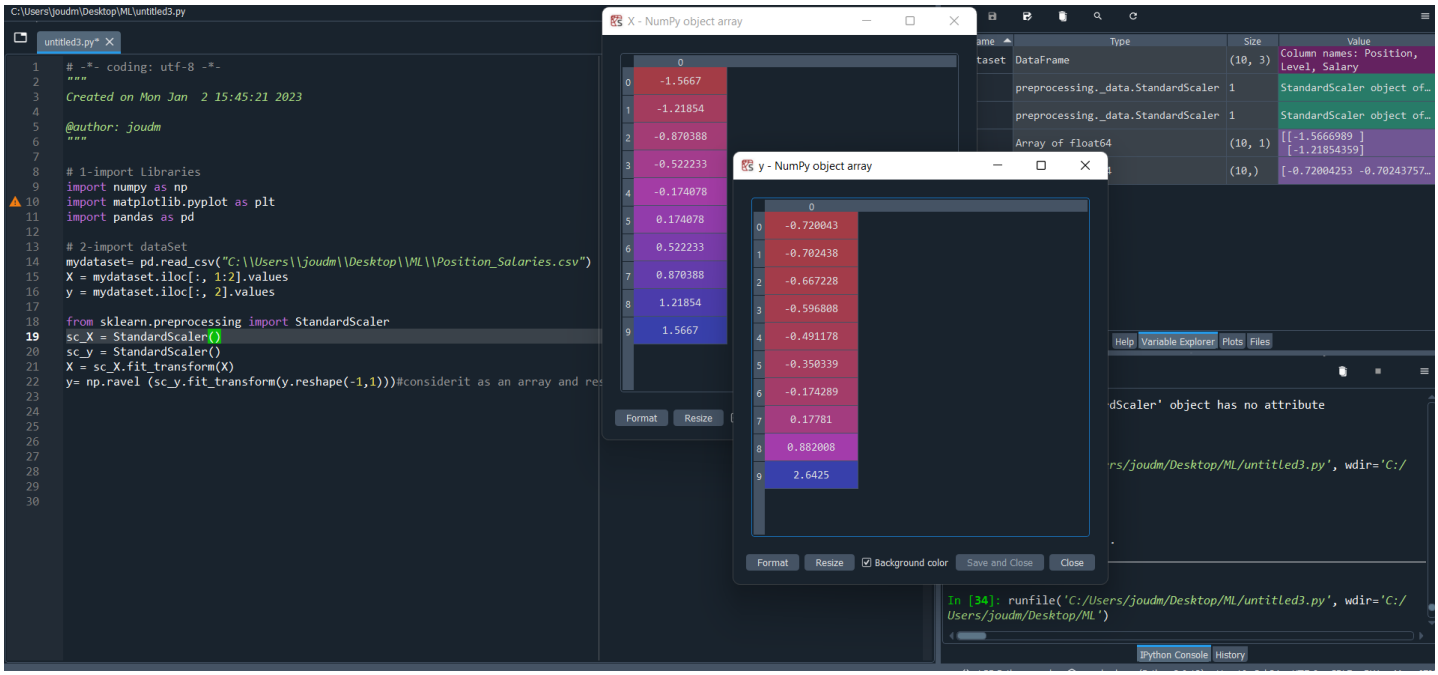
5. Visualize the polynomial linear regression results (show screenshot).

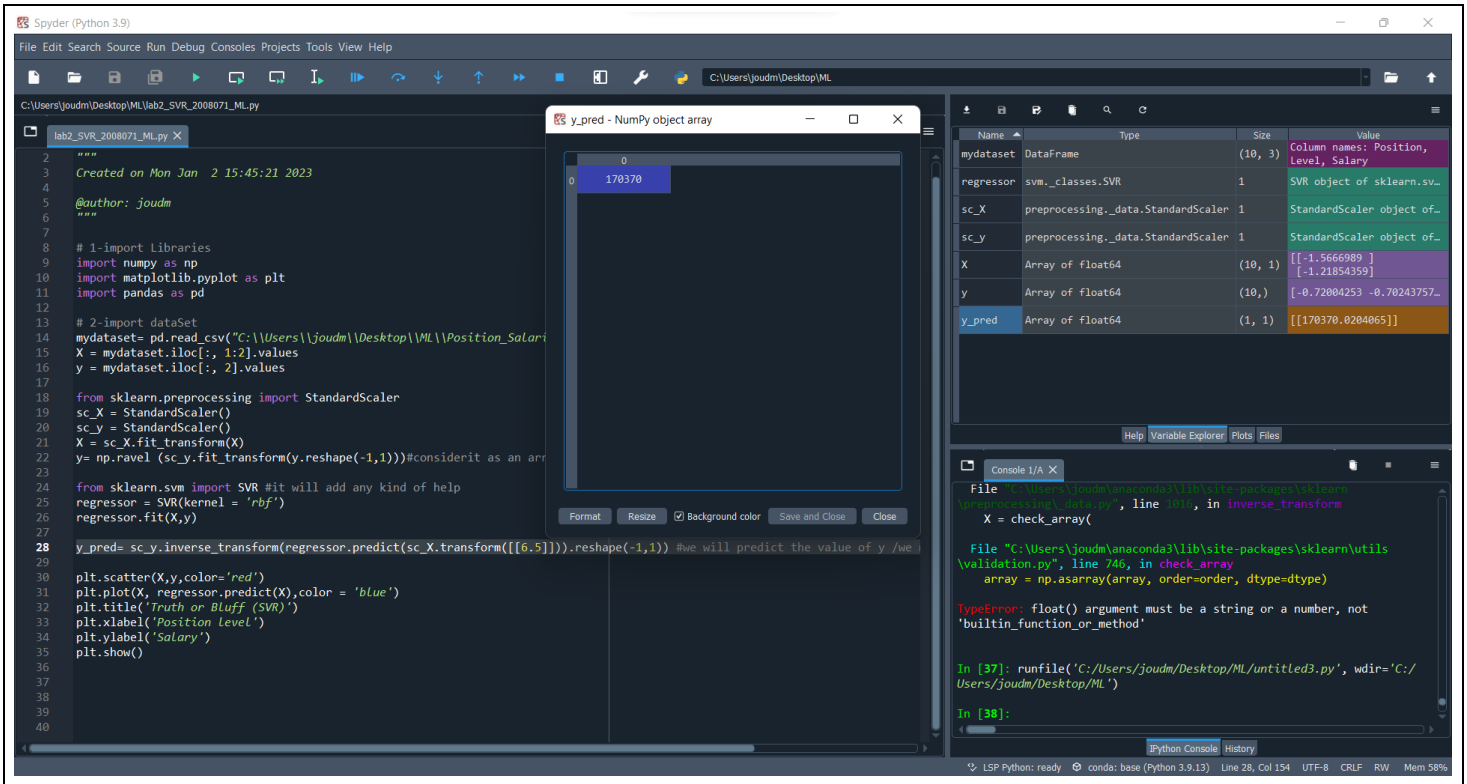


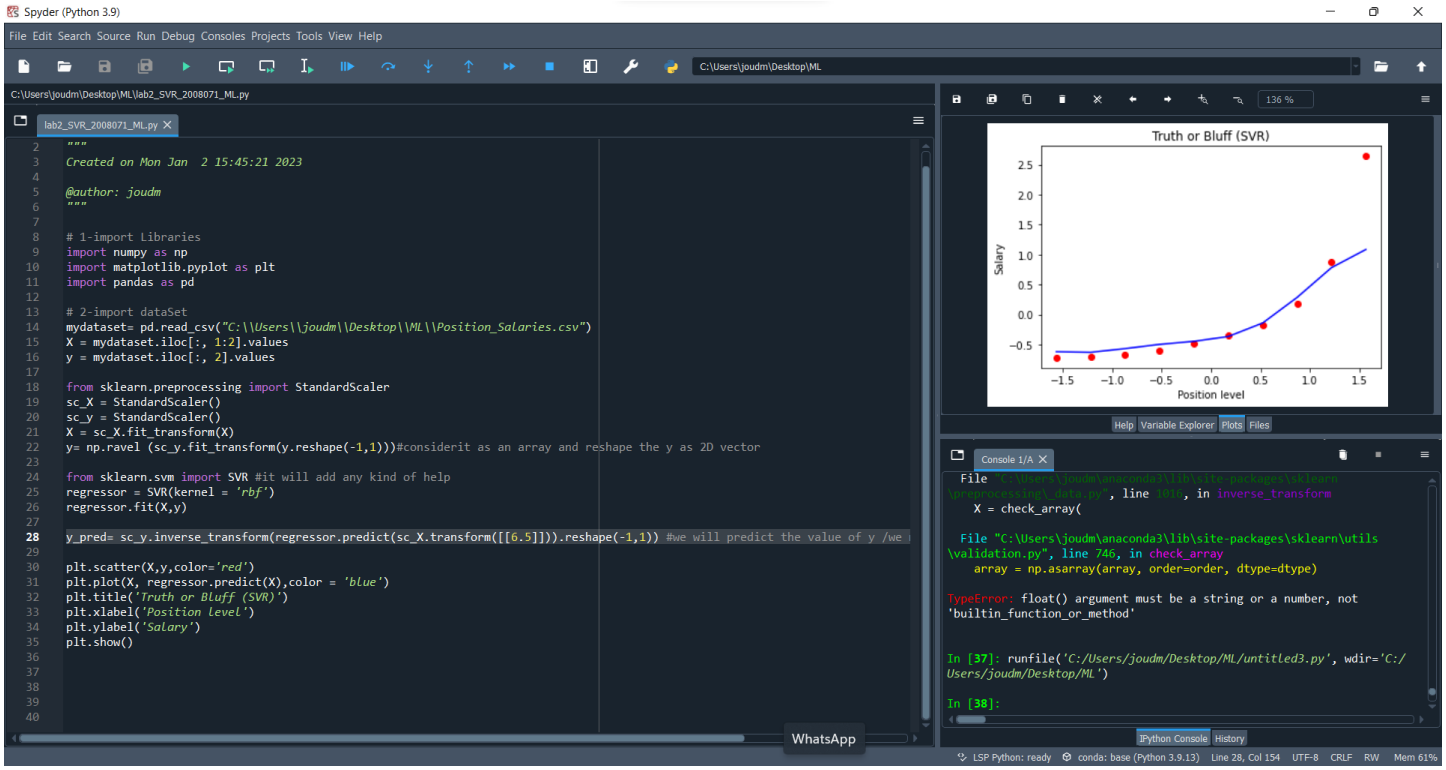
6. Visualize the support vector regression results (show screenshot).



Support Vector Regression (SVR)




Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\joudm\Desktop\ML
C:\Users\joudm\Desktop\ML\lab2_SVR_2008071_ML.py
lab2_SVR_2008071_ML.py
"""
Created on Mon Jan 2 15:45:21 2023
@author: joudm
"""
# 1-import Libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
# 2-import dataSet
mydataset= pd.read_csv("C:\\Users\\joudm\\Desktop\\ML\\Position_Salari
X = mydataset.iloc[:, 1:2].values
y = mydataset.iloc[:, 2].values
from sklearn.preprocessing import StandardScaler
sc_X = StandardScaler()
sc_y = StandardScaler()
X = sc_X.fit_transform(X)
y= np.ravel (sc_y.fit_transform(y.reshape(-1,1)))#considerit as an arr
from sklearn.svm import SVR #it will add any kind of help
regressor = SVR(kernel = 'rbf')
regressor.fit(X,y)
y_pred= sc_y.inverse_transform(regressor.predict(sc_X.transform([[6.5]])).reshape(-1,1)) #we will predict the value of y /we
plt.scatter(X,y,color='red')
plt.plot(X, regressor.predict(X),color = 'blue')
plt.title('Truth or Bluff (SVR)')
plt.xlabel('Position level')
plt.ylabel('Salary')
plt.show()
y_pred - NumPy object array
0
0 170370
Format Resize Background color Save and Close Close
Name Type Size Value
mydataset DataFrame (10, 3) Column names: Position, Level, Salary
regressor svm._classes.SVR 1 SVR object of sklearn.sv…
sc_X preprocessing._data.StandardScaler 1 StandardScaler object of…
sc_y preprocessing._data.StandardScaler 1 StandardScaler object of…
X Array of float64 (10, 1) [[-1.5666989] [-1.21854359]
y Array of float64 (10,) [-0.72004253 -0.70243757…
y_pred Array of float64 (1, 1) [[170370.0204065]]
Help Variable Explorer Plots Files
Console 1/A
File "C:\Users\joudm\anaconda3\lib\site-packages\sklearn\preprocessing_data.py", line 1011, in inverse_transform
X = check_array(
File "C:\Users\joudm\anaconda3\lib\site-packages\sklearn\utils\validation.py", line 746, in check_array
array = np.asarray(array, order=order, dtype=dtype)
TypeError: float() argument must be a string or a number, not 'builtin_function_or_method'
In [37]: runfile('C:/Users/joudm/Desktop/ML/untitled3.py', wdir='C:/Users/joudm/Desktop/ML')
In [38]:
IPython Console History
LSP Python: ready conda: base (Python 3.9.13) Line 28, Col 154 UTF-8 CRLF RW Mem 58%


Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\joudm\Desktop\ML
C:\Users\joudm\Desktop\ML\lab2_SVR_2008071_ML.py
lab2_SVR_2008071_ML.py
"""
Created on Mon Jan 2 15:45:21 2023
@author: joudm
"""
# 1-import Libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
# 2-import dataSet
mydataset= pd.read_csv("C:\\Users\\joudm\\Desktop\\ML\\Position_Salaries.csv")
X = mydataset.iloc[:, 1:2].values
y = mydataset.iloc[:, 2].values
from sklearn.preprocessing import StandardScaler
sc_X = StandardScaler()
sc_y = StandardScaler()
X = sc_X.fit_transform(X)
y= np.ravel (sc_y.fit_transform(y.reshape(-1,1)))#considerit as an array and reshape the y as 2D vector
from sklearn.svm import SVR #it will add any kind of help
regressor = SVR(kernel = 'rbf')
regressor.fit(X,y)
y_pred= sc_y.inverse_transform(regressor.predict(sc_X.transform([[6.5]])).reshape(-1,1)) #we will predict the value of y /we
plt.scatter(X,y,color='red')
plt.plot(X, regressor.predict(X),color = 'blue')
plt.title('Truth or Bluff (SVR)')
plt.xlabel('Position level')
plt.ylabel('Salary')
plt.show()
Truth or Bluff (SVR)
Salary
Position level
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Help Variable Explorer Plots Files
Console 1/A
File "C:\Users\joudm\anaconda3\lib\site-packages\sklearn\preprocessing_data.py", line 1011, in inverse_transform
X = check_array(
File "C:\Users\joudm\anaconda3\lib\site-packages\sklearn\utils\validation.py", line 746, in check_array
array = np.asarray(array, order=order, dtype=dtype)
TypeError: float() argument must be a string or a number, not 'builtin_function_or_method'
In [37]: runfile('C:/Users/joudm/Desktop/ML/untitled3.py', wdir='C:/Users/joudm/Desktop/ML')
In [38]:
IPython Console History
WhatsApp
LSP Python: ready conda: base (Python 3.9.13) Line 28, Col 154 UTF-8 CRLF RW Mem 61%

7. Based on the scenario question, write your opinion about every model’s result.

A linear regression line does not easily fit the first graph. In order to achieve a minimum value of error and obtain the best estimate of salary based on the employee's entire position level, we used polynomial regression to fit polynomial line at the second graph (5.5)

The Polynomial linear regression module provides the best result of salary, which is as the two modules demonstrate, where (132148.43750003).

We can determine that the new employee deserves $133,000 based on the module's outcome.

And I have notice that the SVR is sensitive to outlairs, it does not care about the data outside the margin,it try to find the line that keeps as many points as possible within epsilon. But outliers will still affect.

Good Luck ☺