

COGNITIVE ANALYSIS LAB

Submitted by: Submitted to:

Abhishek Chauhan Sugandha Sharma

500096952

B2(H)

Index

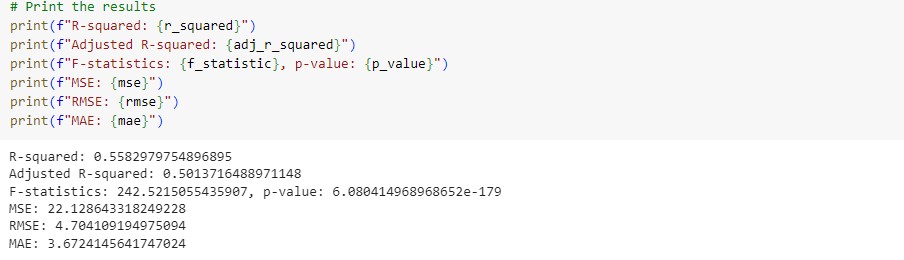
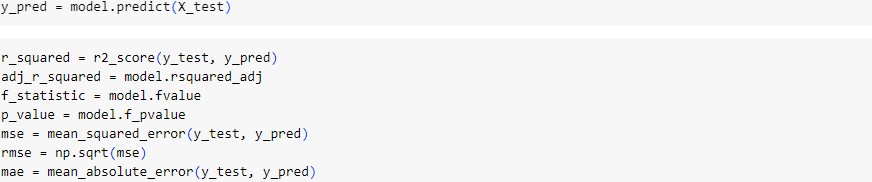
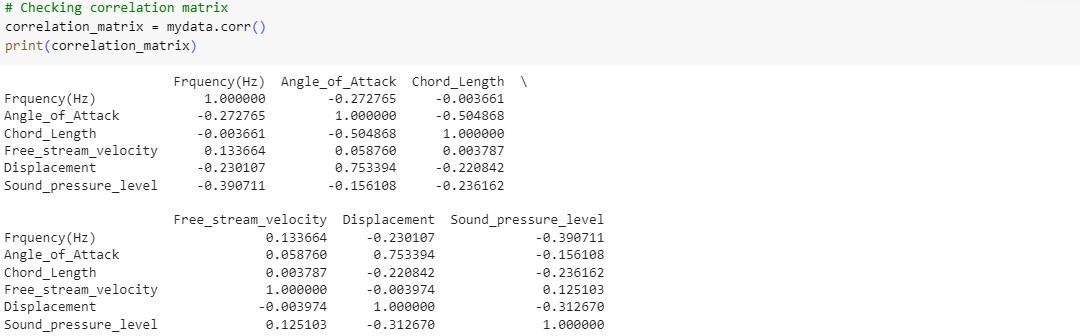
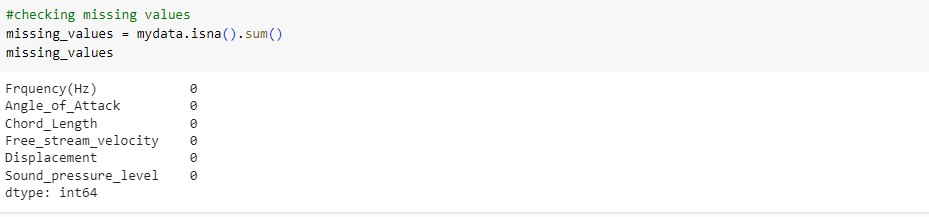
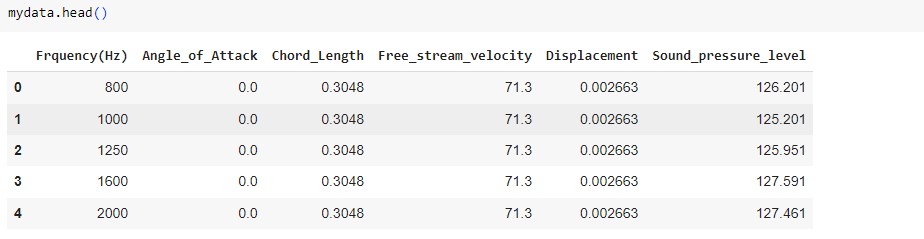
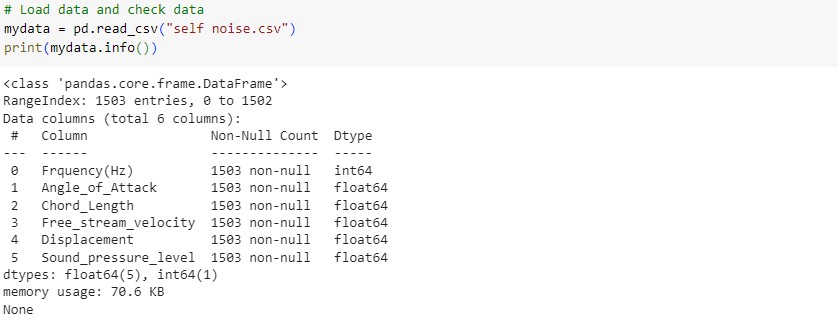
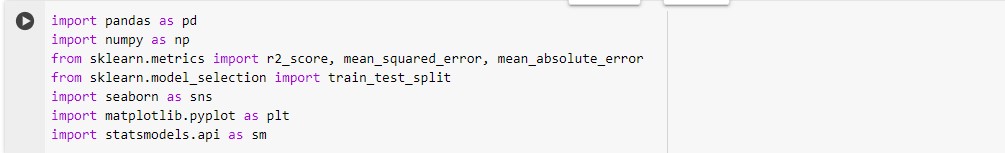
|  |  |  |  |
| --- | --- | --- | --- |
| Sno. | Name | Date | Page |
| 1. | Analysing the dataset | 17-01-2024 | 3 |
| 2. | Ridge and lasso algorithm | 24-01-2024 | 6 |
| 3. | Logistic and Polynomial Regression | 31-01-2024 | 11 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

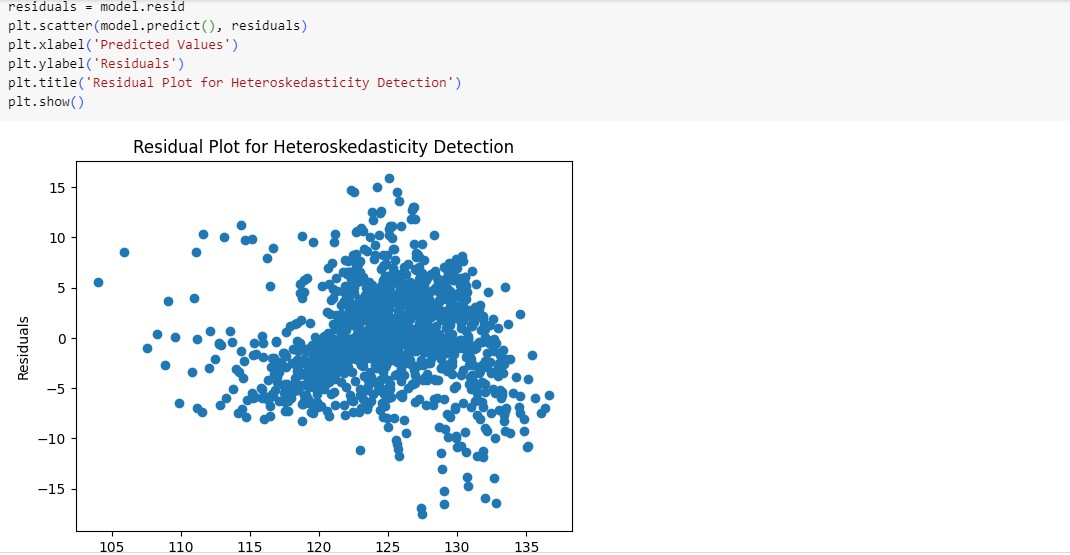
Lab Assignment – 1

Implementing Multilinear and linear regression

Dataset:- <http://blog.hackerearth.com/wp-content/uploads/2016/12/airfoil_self_noise.csv>

Code:-





Interpretation:-

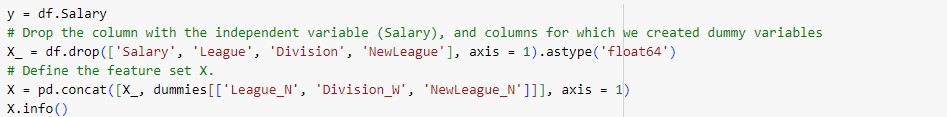
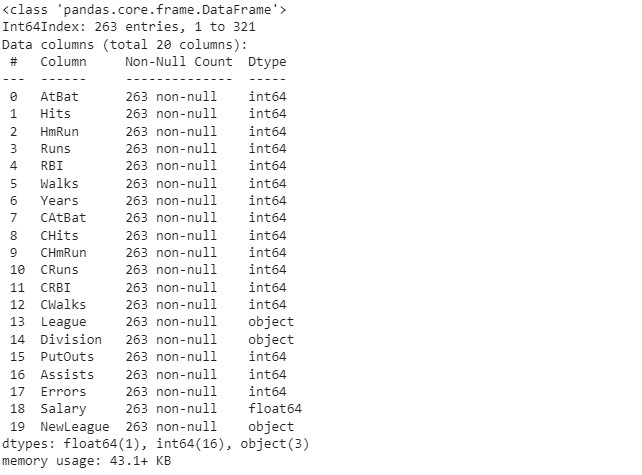
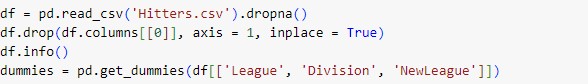
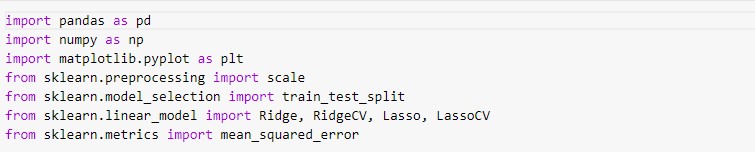
The model explains approximately 55.8% of the variability in the dependent variable. The F-statistic is highly significant (p < 0.001), indicating that the model as a whole is significant. The mean squared error (MSE) is 22.13, and the root mean squared error (RMSE) is 4.70, suggesting the model's predictions are reasonably close to the actual values on average. The mean absolute error (MAE) is 3.67, indicating the average absolute difference between predicted and actual values. Overall, the model demonstrates good explanatory power and predictive accuracy, though further validation may be necessary.

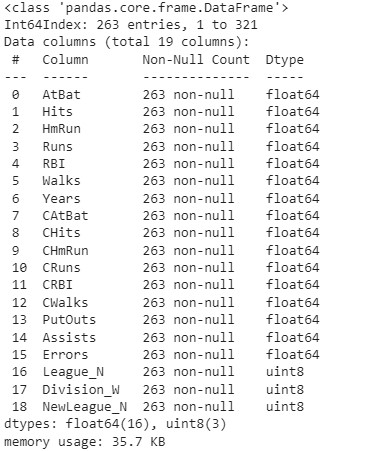
Lab Assignment – 2

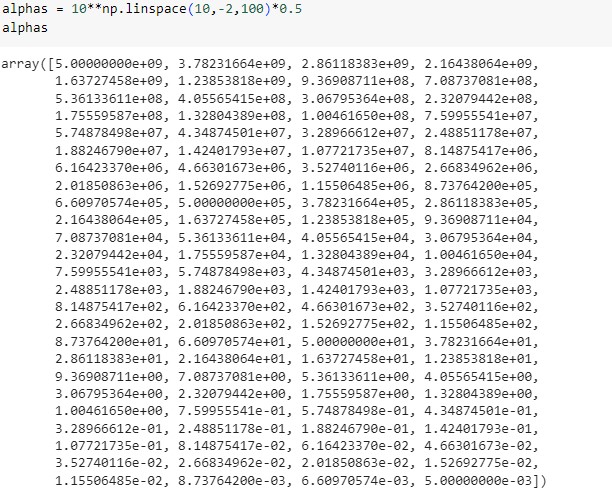
Implement Ridge Regularization and the Lasso Regularization in Python

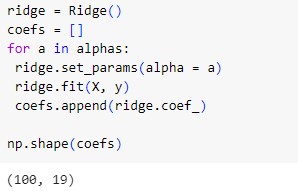
Dataset: <https://github.com/JWarmenhoven/ISLRpython/blob/master/Notebooks/Data/Hitters.csv>

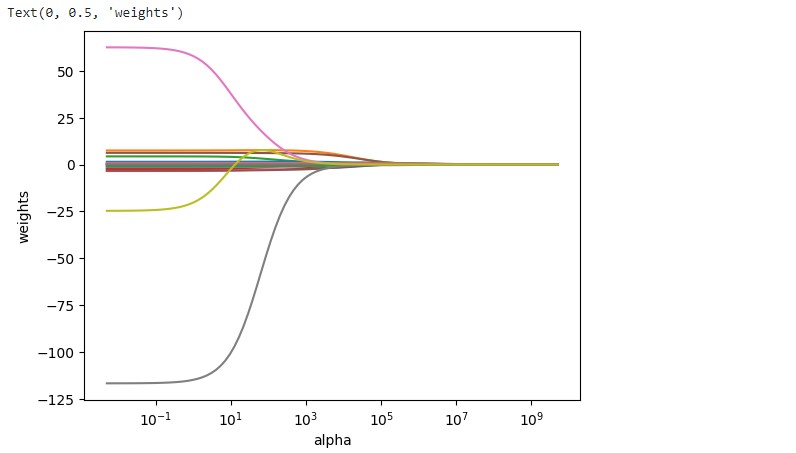
Code:-

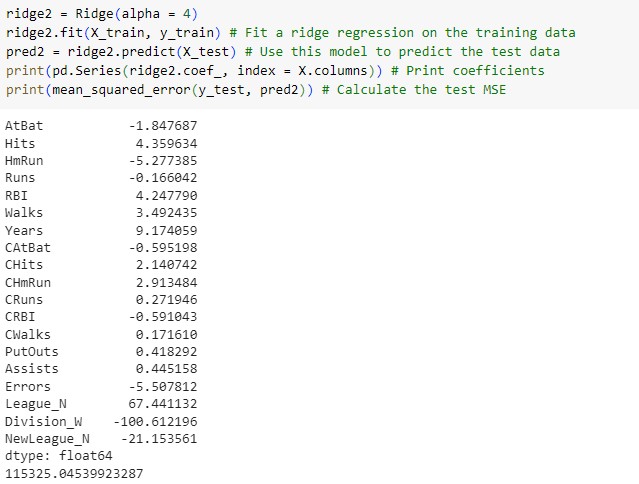


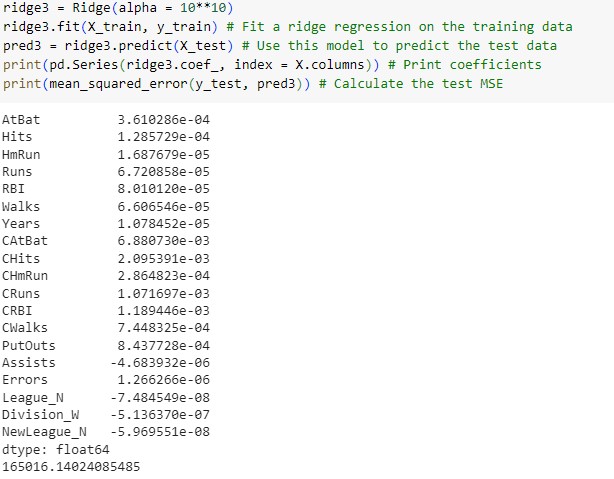


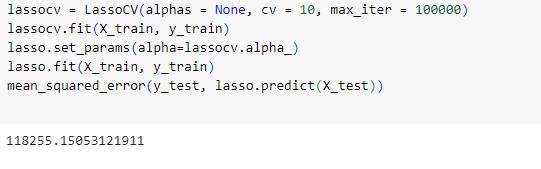
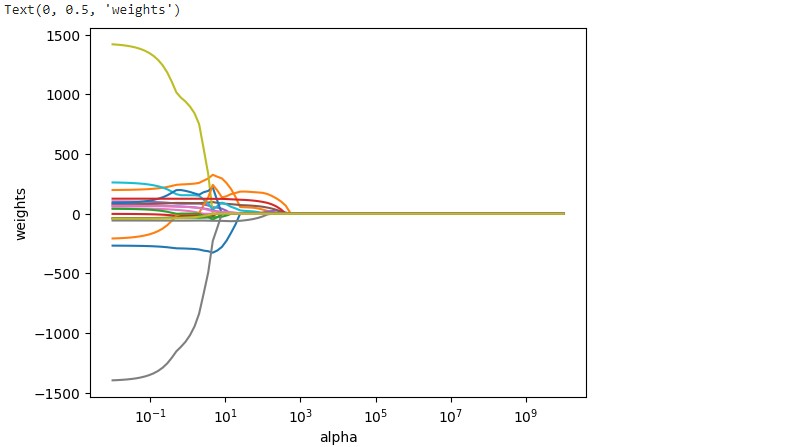
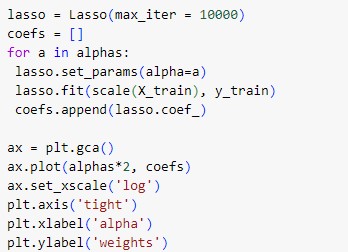
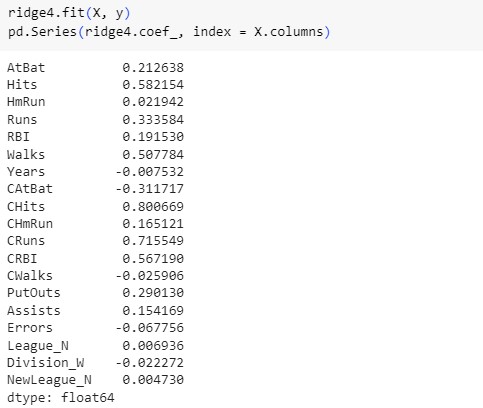
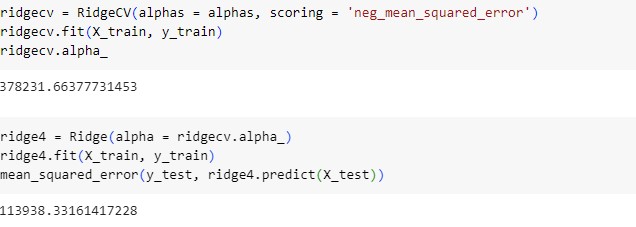
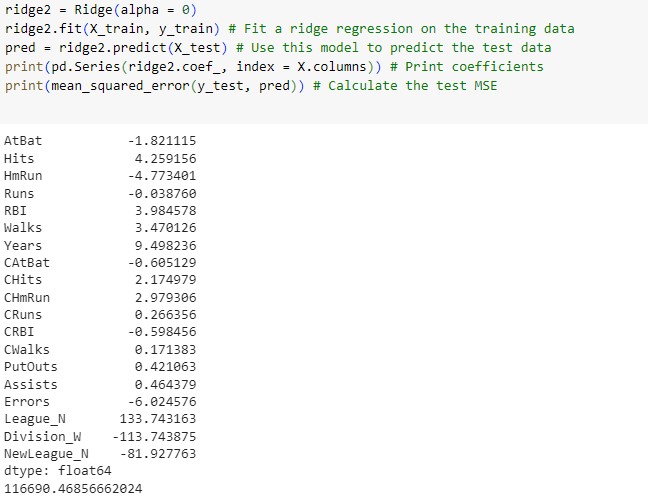


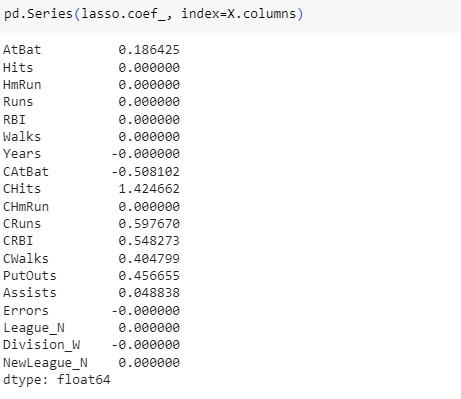












**Interpretation:**

* Ridge and lasso regression are regularization techniques used to prevent overfitting in linear regression models.
* The choice of alpha determines the strength of regularization, with higher alpha leading to more regularization.
  + As we can see from the output of the code, when the value of alpha is small the output values are significantly small but when the alpha is increased the output values decreases with higher rate.
* The code assesses model performance using mean squared error on the test set for both ridge and lasso regression.
* Cross-validation is employed to find optimal alpha values for ridge and lasso, enhancing model generalization.

\

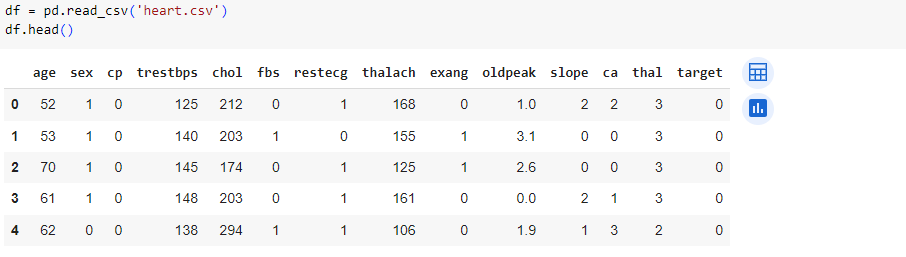
Lab Assignment – 3

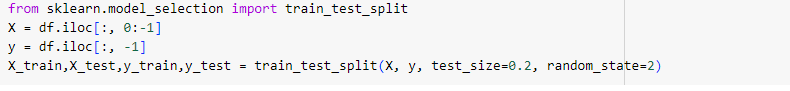
Implement Logistic Regression and Polynomial Regression

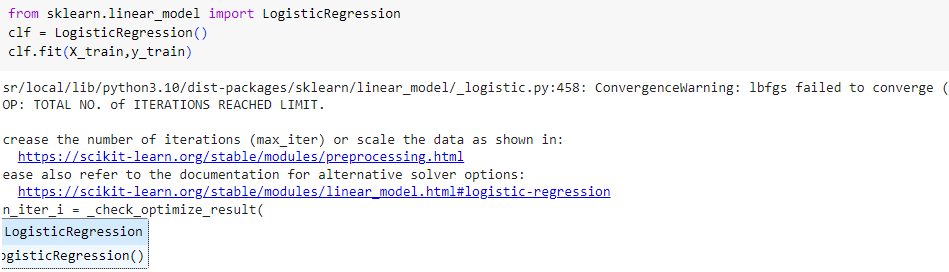
Dataset: <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset?resource=download>

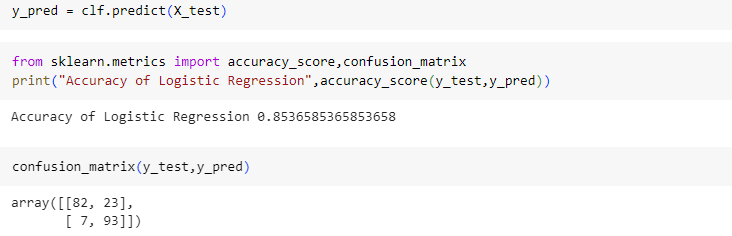
Code:-

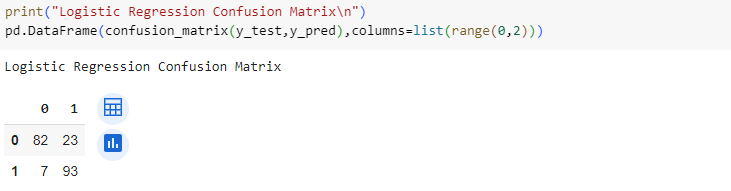










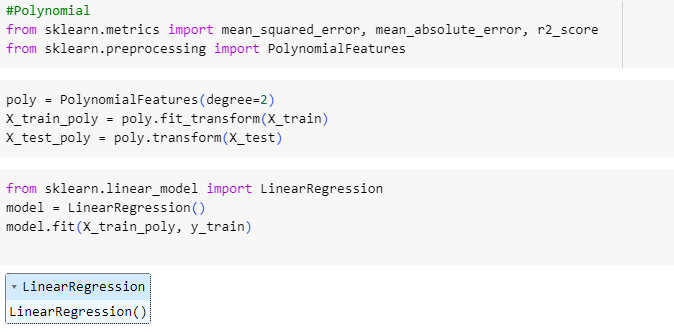


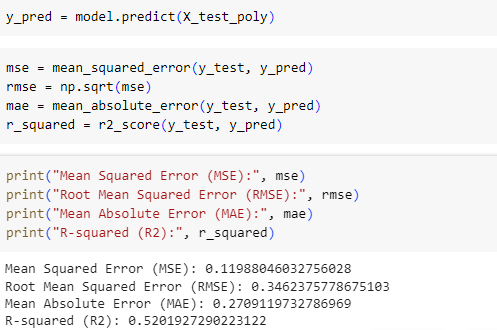
Interpretation:-

1. The confusion matrix provides a breakdown of the model's predictions versus the actual values. In this case:

* True Positives (TP): 82 - The model correctly predicted 82 instances of the positive class.
* False Positives (FP): 23 - The model incorrectly predicted 23 instances as positive when they were actually negative.
* False Negatives (FN): 7 - The model incorrectly predicted 7 instances as negative when they were actually positive.
* True Negatives (TN): 93 - The model correctly predicted 93 instances of the negative class.

1. The precision of 0.80 indicates that when the model predicts a positive outcome, it is correct 80% of the time.
2. The recall of 0.93 indicates that the model captures 93% of all actual positive instances.
3. The F1 score of 0.86 suggests that the model achieves a good balance between precision and recall.





Interpretation:-

* The relatively low R-squared value (0.5202) suggests that the polynomial regression model explains only about 52.02% of the variance in the target variable. This indicates that there may be other factors not captured by the model that influence the target variable.
* The MSE, RMSE, and MAE values provide insights into the accuracy of the model's predictions. The lower these values, the better the model performance. In this case, the values are relatively low, indicating that the model's predictions are reasonably close to the actual values.