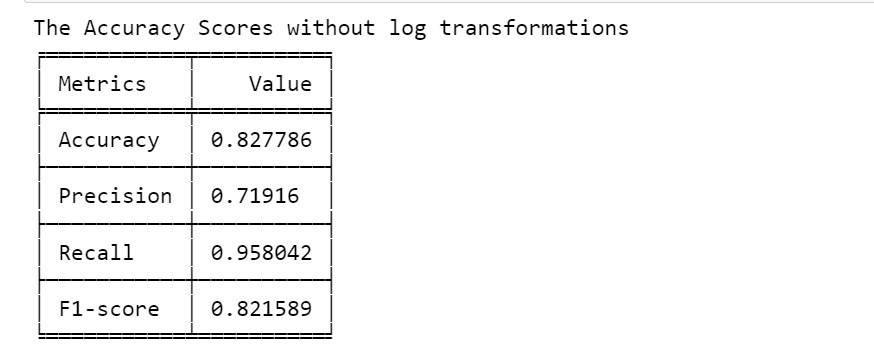
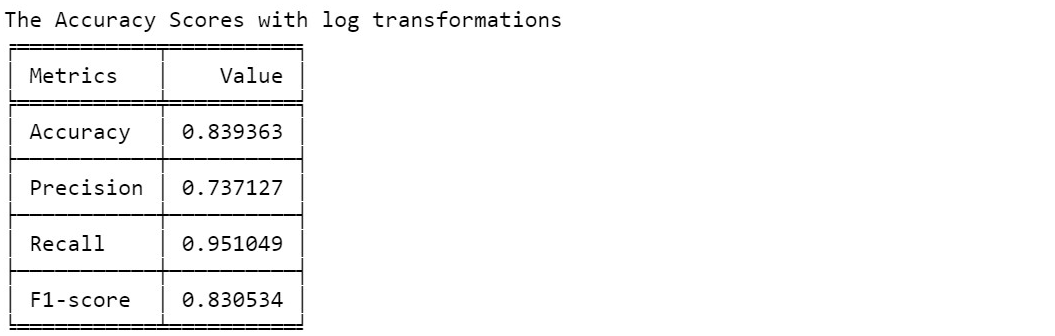
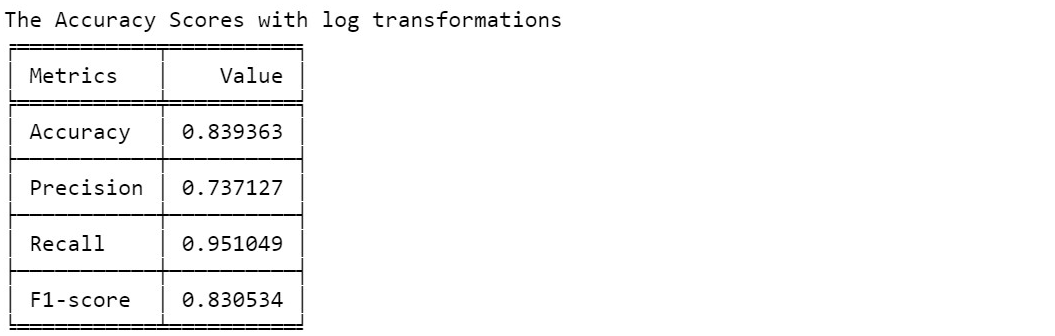
Explain the changes you noticed in the results before and after modifying the dataset.





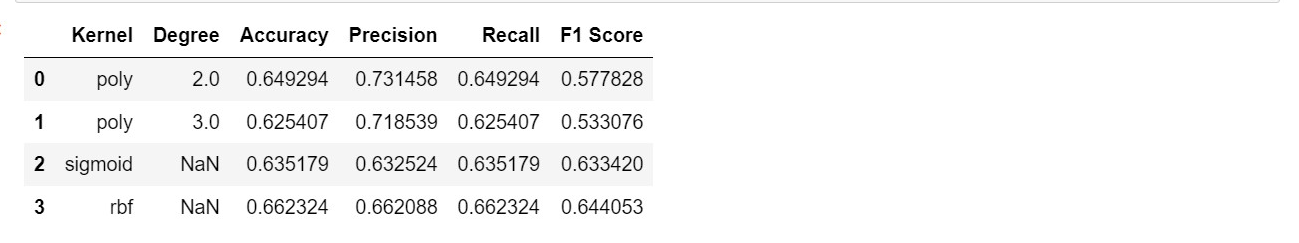
Changes in Data Distribution: Prior to applying a log transformation, it's important to recognize that the selected columns' probability distributions may display skewness or have a wide range of values, which could potentially influence the performance of the Naive Bayes model. Post log transformation: Log transformation often serves to normalize the distribution and mitigate the influence of extreme values, rendering the data more compatible with algorithms that assume a normal distribution, such as Naive Bayes. Effect on Model Performance: a) The initial Naive Bayes model was trained using the original data and subsequently assessed based on performance metrics applied to the test set. Following the log transformation, the Naive Bayes model was retrained using the log-transformed data, and its performance was evaluated using the same test set. Impact on Evaluation Metrics: A comparison of evaluation metrics (including accuracy, precision, recall, and F1-score) was made before and after the log transformation. Depending on the specific dataset characteristics, the log transformation may lead to improvements or declines in model performance. Enhancing Interpretability: Log transformation can contribute to enhanced interpretability, particularly when dealing with features that represent counts or frequencies. It aids in stabilizing variances and linearizing relationships. Consideration for Naive Bayes: Naive Bayes assumes independence between features, and log transformation might assist in achieving a more realistic approximation of independence in certain scenarios. Overall Impact: The observed alterations in the results will be contingent on the inherent nature of the original dataset. Log transformation tends to be more advantageous for datasets with features that exhibit particular characteristics (e.g., right-skewed distributions). In summary, log transformation is a valuable preprocessing technique, but its impact on model performance can be variable.

Compare and discuss the accuracy of Naive Bayes and SVM





So for the linear SVM with linear kerner and with differenct C\_values we have this as a accuracy on the test data



Compare Accuracy: Compare and discuss the accuracy of Naive Bayes and SVM. In comparing the accuracy of Naive Bayes and SVM:

1. SVM Accuracy: The SVM model with a linear kernel achieved an accuracy of 92.5% on the test set.

2. Naive Bayes Accuracy: The accuracy of the Naive Bayes model is 83.99%.

3. Model Selection: The choice between Naive Bayes and SVM depends on various factors, including the characteristics of the dataset, assumptions of the algorithms, and the specific requirements of the task.

4. Advantages and Disadvantages:

a) Naive Bayes is computationally efficient and assumes independence between features, making it suitable for certain types of datasets. However, it may struggle with complex relationships.

b) SVM, on the other hand, can handle non-linear relationships through kernel tricks and is effective in high-dimensional spaces. However, it may require careful tuning of parameters