Kean Gabriel E. Salvahan

BSCS – 3B

Interpretation & Discussion

Write a short report addressing the following:

• **What do the results of the confusion matrix indicate?**

The confusion matrix shows that the Logistic regression model correctly classified the vast majority of instances in both classes. With only 2 false positives and 1 false negative out of 114 test samples, the model demonstrates high reliability.

• **How consistent is the model's performance based on 5-Fold Cross Validation?**

The 5-Fold Cross-Validation yielded a mean accuracy of approximately 0.97 with a very low standard deviation. This indicates that the model's performance is stable and consistent across different subsets of the data. There is no significant overfitting to a particular train-test split. The model generalizes well to unseen data as confirmed by similar scores across all five folds.

• **What insights can be derived from the learning curve?**

The learning curve reveals two key insights which are the training accuracy remains high across all training set sizes while validation accuracy starts lower but improves steadily as more data is used. The gap between training and validation scores narrows as the dataset grows but a small gap remains even at full size. This pattern indicates mild overfitting especially when training on smaller datasets. However, since the validation score continues to rise and stabilizes near 0.8, the model benefits from additional data and is not underfitting.

• **How can the model be improved?**

Although the model already performs well, there can be several strategies that could enhance it further like Hyperparameter tuning, Class balancing and collecting more data. As suggested by the learning curve, additional high quality samples may close the remaining generalization gap.