Machine Problem 2			
Topic:	Evaluating Machine Learning Model Performance	**Course Code:**	CSST102
Course Title:	Basic Machine Learning	**Academic Year:**	2025-2026
Student Name	Macatangay, Shaine Carla P.	**Section:**	BSCS 3A

Machine Problem No. 2: Evaluating Machine Learning Model Performance

This report presents the results and analysis of applying Logistic Regression on the Iris dataset to evaluate machine learning model performance. The activity follows a complete data science workflow data preparation, model training, cross-validation, and evaluation to assess accuracy, stability, and generalization. The experiment includes visualizations such as the confusion matrix and learning curve, as well as a comparison with Decision Tree and SVM models for deeper insight.

Interpretation & Discussion

1. What do the results of the confusion matrix indicate?

The confusion matrix shows that the Logistic Regression model correctly classified almost all samples of the three Iris species. Most predictions fall along the diagonal line of the matrix, which means the model has learned to distinguish between setosa, versicolor, and virginica effectively. The minimal number of misclassifications indicates that feature overlap among the classes is low and the decision boundaries are well established. This result reflects a strong model with high predictive reliability and little confusion between categories. In practical terms, it means the algorithm has accurately captured the patterns in the dataset. Therefore, the confusion matrix confirms that the model performs with excellent precision and accuracy across all classes.

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

The 5-Fold Cross Validation produced accuracy scores ranging between approximately 96% and 98%, showing very little variation between folds. The small standard deviation indicates that the model performs consistently across different subsets of the dataset. This consistency implies that the Logistic Regression model generalizes well and does not depend heavily on any specific portion of the data. It demonstrates the model's stability and reliability when exposed to unseen samples. The close similarity between training and testing results also suggests minimal variance error. Overall, the cross-validation results confirm that the model's performance is both strong and consistent across multiple trials.

3. What insights can be derived from the learning curve?

The learning curve shows that both training and validation accuracies converge near 97% as the number of training samples increases. This convergence pattern indicates that the model is neither overfitting nor underfitting the data. The absence of a large gap between the two curves shows that the model has learned generalizable patterns instead of memorizing training examples. As the dataset grows, accuracy stabilizes, suggesting that the current dataset size is already sufficient for optimal performance. This balance also reflects good model complexity and appropriate regularization. Hence, the learning curve demonstrates that the Logistic Regression model is well-fitted and capable of maintaining stable accuracy with more data.

4. How can the model be improved?

Although the model already achieves high accuracy, further improvements can enhance its robustness and adaptability. One approach is to use GridSearchCV to optimize hyperparameters such as C (regularization strength) and the solver type. Another method is to include polynomial features to capture possible non-linear relationships between features. Using a larger and more diverse dataset could also test the model's performance on broader conditions and reduce dataset bias. Comparing Logistic Regression with non-linear classifiers like SVM or Random Forest can reveal additional insights about model behavior. Overall, targeted hyperparameter tuning and model comparison can help achieve even higher generalization and predictive accuracy.

Performance Summary

The Logistic Regression model achieved excellent predictive performance on the Iris dataset, with an overall testing accuracy of approximately 97%. Both training and testing results showed minimal difference, proving that the model learned meaningful patterns without overfitting. Evaluation metrics such as precision, recall, and F1-score were all consistently high, confirming that predictions across all three Iris species were reliable and balanced. The 5-Fold Cross Validation results further supported this consistency, yielding an average accuracy of around 96–98% with a low standard deviation. The learning curve visualization showed that training and validation accuracies converge, indicating a well-fitted model capable of generalizing effectively. Overall, the model demonstrates strong performance, stability, and suitability for multi-class classification tasks.

Conclusion

In conclusion, the Logistic Regression model performed exceptionally well on the Iris dataset, achieving high accuracy, precision, recall, and F1-scores across all classes. The results of the confusion matrix confirm that the model effectively differentiates between the three Iris species with minimal misclassifications. The 5-Fold Cross Validation further validates the model's consistency, showing stable performance and strong generalization across multiple data partitions. The learning curve indicates that the model is well-fitted, maintaining balanced accuracy between training and validation sets without signs of overfitting or underfitting. While the model already demonstrates reliability, it can still be improved through hyperparameter tuning, feature engineering, and testing on larger or more complex datasets. Overall, this study shows that Logistic Regression is a powerful, interpretable, and efficient algorithm for multi-class classification problems in machine learning.