Assignment 1: Evaluate Regression and Classifier Metrics

Anuj Bansal

Data Sciences, National University

DDS-8555 v1: Predictive Analysis

Dr. Mohammad

July 20th, 2025

This week’s report goes over regression and classifier model. Regression in machine learning refers to a [supervised learning](https://www.geeksforgeeks.org/machine-learning/supervised-machine-learning/)technique where the goal is to predict a continuous numerical value based on one or more independent features (Regression in machine learning, 2025).  The assignment builds two different regression models to predict sepal width; first by mean of petal length and the second using sepal length minus petal width. The first model (est1, mean of petal length) performed better as a regressor because it has lower MAE, MSE, and MAPE values (scikit-learn developers , 2007-2025) compared to the second model. Lower error values indicate better predictive performance. The results are summarized in Appendix A.

Another model included in the report is classifier model.  classification models play a pivotal role. They are the go-to tools for solving problems where the goal is to categorize data into predefined classes or groups (Classification Model , 2023). Two different model are built to evaluate ‘type’ based on sepal length. These models are compared for accuracy, precision, recall and F1 score. The first classifier uses quantiles from est3 (25th and 50th percentiles of sepal length). The second classifier uses quantiles from est4 (50th and 75th percentiles of sepal length). The first classifier (using est3 quantiles) has an accuracy of 0.20 (20%). The **second classifier** (using est4 quantiles) has an accuracy of **0.07** (7%). Other metrics (precision, recall, f1-score) are also higher for the first classifier.

We finish this report with a few ways we can improve these models and will learn about some of these techniques as the course progresses (RITHP, 2023)

1. **Use More Features:**   
   Current classifiers and regressors use only simple rules or single features. We can try using all available features as input to the models.
2. **Apply Machine Learning Algorithms:**  
   by using established algorithms such as logistic regression, decision trees, random forests, or support vector machines for classification, and linear regression or more advanced regressors for regression tasks.
3. **Feature Engineering:**  
   Creating new features or transform existing ones (e.g., polynomial features, interaction terms, normalization/scaling).
4. **Hyperparameter Tuning:**  
   Use cross-validation and grid search to find the best parameters for these models.
5. **Model Evaluation:**  
   Use proper train/test splits and cross-validation to evaluate model performance more reliably.
6. **Handle Class Imbalance:**  
   If classes are imbalanced, considering techniques like oversampling, undersampling, or using class weights.
7. **Visualize Data:**  
   Use plots to understand feature distributions and relationships, which can guide feature selection and engineering.

Appendix A

Results for regression model:

* For the first model (est1):
  + MAE: 0.694
  + MSE: 0.602
  + MAPE: 0.242
* For the second model (est2):
  + MAE: 1.543
  + MSE: 2.526
  + MAPE: 0.522

Results of Classification model

Model #1

precision recall f1-score support

0 0.00 0.00 0.00 0

1 0.00 0.00 0.00 7

2 0.35 0.86 0.50 7

3 0.00 0.00 0.00 6

42 0.00 0.00 0.00 1

50 0.00 0.00 0.00 1

53 0.00 0.00 0.00 1

54 0.00 0.00 0.00 1

56 0.00 0.00 0.00 1

58 0.00 0.00 0.00 1

63 0.00 0.00 0.00 1

66 0.00 0.00 0.00 1

82 0.00 0.00 0.00 1

96 0.00 0.00 0.00 1

accuracy 0.20 30

macro avg 0.03 0.06 0.04 30

weighted avg 0.08 0.20 0.12 30

Model #2

precision recall f1-score support

0 0.00 0.00 0.00 0

1 0.12 0.14 0.13 7

2 0.11 0.14 0.12 7

3 0.00 0.00 0.00 6

42 0.00 0.00 0.00 1

50 0.00 0.00 0.00 1

53 0.00 0.00 0.00 1

54 0.00 0.00 0.00 1

56 0.00 0.00 0.00 1

58 0.00 0.00 0.00 1

63 0.00 0.00 0.00 1

66 0.00 0.00 0.00 1

82 0.00 0.00 0.00 1

96 0.00 0.00 0.00 1

accuracy 0.07 30

macro avg 0.02 0.02 0.02 30

weighted avg 0.06 0.07 0.06 30

# References

*Classification Model* . (2023, sep 30). Retrieved from medium: https://medium.com/@karan.kamat1406/which-classification-model-should-you-use-a-cheat-sheet-for-machine-learning-practitioners-3fea0bcab04e

*Regression in machine learning*. (2025, Jan 13). Retrieved from geek for geeks: https://www.geeksforgeeks.org/machine-learning/regression-in-machine-learning/

RITHP. (2023, jan 4). *Medium*. Retrieved from Improving Precision and Recall in Machine Learning: Tips and Techniques: https://medium.com/@rithpansanga/improving-precision-and-recall-in-machine-learning-tips-and-techniques-acb5a5fd27a6

scikit-learn developers . (2007-2025). *Metrics and scoring: quantifying the quality of predictions*. Retrieved from scikit-learn: https://scikit-learn.org/stable/modules/model\_evaluation.html#classification-metrics